

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:24:12 ; Search time 116.27 Seconds

Title: US-09-966-608-2

Sequence: 1 DLASAVGIQSGSIFHFKSK.....SLSAEGQAHVLLALRDVEQI 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main: \*

1:	/cgn2.6/p/odata/2/paa/PCrUIS.COMB.pep.*
2:	/cgn2.6/p/odata/2/paa/US06.COMB.pep.*
3:	/cgn2.6/p/odata/2/paa/US07.COMB.pep.*
4:	/cgn2.6/p/odata/2/paa/US08.COMB.pep.*
5:	/cgn2.6/p/odata/2/paa/US081.COMB.pep.*
6:	/cgn2.6/p/odata/2/paa/US082.COMB.pep.*
7:	/cgn2.6/p/odata/2/paa/US083.COMB.pep.*
8:	/cgn2.6/p/odata/2/paa/US084.COMB.pep.*
9:	/cgn2.6/p/odata/2/paa/US085.COMB.pep.*
10:	/cgn2.6/p/odata/2/paa/US086.COMB.pep.*
11:	/cgn2.6/p/odata/2/paa/US087.COMB.pep.*
12:	/cgn2.6/p/odata/2/paa/US088.COMB.pep.*
13:	/cgn2.6/p/odata/2/paa/US089.COMB.pep.*
14:	/cgn2.6/p/odata/2/paa/US090.COMB.pep.*
15:	/cgn2.6/p/odata/2/paa/US091.COMB.pep.*
16:	/cgn2.6/p/odata/2/paa/US092.COMB.pep.*
17:	/cgn2.6/p/odata/2/paa/US093.COMB.pep.*
18:	/cgn2.6/p/odata/2/paa/US094.COMB.pep.*
19:	/cgn2.6/p/odata/2/paa/US095.COMB.pep.*
20:	/cgn2.6/p/odata/2/paa/US096.COMB.pep.*
21:	/cgn2.6/p/odata/2/paa/US097.COMB.pep.*
22:	/cgn2.6/p/odata/2/paa/US098.COMB.pep.*
23:	/cgn2.6/p/odata/2/paa/US099.COMB.pep.*
24:	/cgn2.6/p/odata/2/paa/US090.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	493	100.0	100	1	PCT-US00-11473-2	Sequence 2, Appli
2	493	100.0	100	1	PCT-US01-16032-3	Sequence 3, Appli
3	493	100.0	100	1	PCT-US01-16032A-3	Sequence 3, Appli
4	493	100.0	100	17	US-09-303-120B-2	Sequence 2, Appli
5	493	100.0	100	19	US-09-575-061-3	Sequence 3, Appli
6	493	100.0	100	22	US-09-820-576-2	Sequence 2, Appli
7	429	87.0	245	16	US-09-325-991A-26292	Sequence 26292, A
8	267	54.2	266	17	US-09-328-352-5572	Sequence 5572, A
9	110.5	22.4	190	1	PCT-US00-11473-6	Sequence 6, Appli

10	110.5	22.4	190	17	US-09-303-1208-6	Sequence 6, April
11	110.5	-22.4	190	22	US-09-820-576-6	Sequence 6, April
12	102.5	20.8	202	23	US-09-902-540-14870	Sequence 14870, A
13	94.5	19.2	200	1	PCT-US00-11473-7	Sequence 7, April
14	94.5	19.2	200	17	US-09-303-1208-7	Sequence 7, April
15	94.5	19.2	200	22	US-09-820-576-7	Sequence 7, April
16	92.5	18.8	193	24	US-60-215-161-4731	Sequence 4731, April
17	92	18.7	193	23	US-09-902-540-16001	Sequence 16001, A
18	90	18.3	222	21	US-09-738-626-4464	Sequence 4464, April
19	89	18.1	192	1	PCT-US00-11473-8	Sequence 8, April
20	89	18.1	192	17	US-09-303-1208-8	Sequence 8, April
21	89	18.1	192	22	US-09-820-576-8	Sequence 8, April
22	88.5	18.0	220	16	US-09-252-691-8207	Sequence 8207, April
23	88.5	18.0	220	16	US-09-252-691-8207	Sequence 8207, April
24	87	17.6	245	23	US-09-902-540-12541	Sequence 12541, A
25	86.5	17.5	177	21	US-09-739-449-8725	Sequence 8725, April
26	86.5	17.5	177	22	US-09-803-110-8725	Sequence 8725, April
27	86	17.4	243	16	US-09-252-691A-26765	Sequence 26765, A
28	85.5	17.3	199	9	US-09-540-336-2051	Sequence 2051, April
29	85.5	17.3	199	24	US-60-128-876-3969	Sequence 3969, April
30	85	17.2	212	16	US-09-352-691A-17137	Sequence 17137, A
31	84	17.0	196	24	US-60-215-61-4814	Sequence 4814, April
32	84	17.0	216	17	US-09-328-352-4476	Sequence 4476, April
33	83.5	16.9	242	19	US-09-543-681A-8329	Sequence 8329, April
34	83	16.8	195	15	US-09-134-001C-3548	Sequence 3548, April
35	83	16.8	195	18	US-09-456-469-4526	Sequence 4526, April
36	81.5	16.5	230	18	US-09-489-0399-13695	Sequence 13695, A
37	81	16.4	190	21	US-09-739-449-9155	Sequence 9155, April
38	81	16.4	190	22	US-09-803-110-9155	Sequence 9155, April
39	81	16.4	244	17	US-09-328-352-5597	Sequence 5597, April
40	80	16.2	216	16	US-09-252-691A-20124	Sequence 20124, April
41	80	16.2	217	16	US-09-252-691A-18564	Sequence 18564, April
42	79.5	16.1	183	24	US-60-215-161-5762	Sequence 5762, April
43	79.5	16.1	193	24	US-60-215-161-5915	Sequence 5915, April
44	79.5	16.1	249	23	US-09-902-540-9786	Sequence 9786, April
45	78.5	15.9	192	16	US-09-252-691A-20144	Sequence 20144, April

## ALIGNMENTS

```

1 RESULT
PCT-US00-11473-2
Sequence 2, Application PC/TUSO011473
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: IBD-Associated Microbial Antigens and Methods of Using
TITLE OF INVENTION: Same
FILE REFERENCE: EP-PM 4142
CURRENT APPLICATION NUMBER: PCT/US00/11473
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 09/303,120
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 100
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Microbial
PCT-US00-11473-2

```

Oy	61	LQSIMGSGEAMAVLVYEMRSLAEGQHVLALRDVYEI	1000
Dd	61	LQSIMGSGEAMAVLVYEMRSLAEGQHVLALRDVYEI	1000

## 2

```

PCT-US01-16032-3
: Sequence 3, Application PC/UTS0116032
: GENERAL INFORMATION:
: APPLICANT: Cedars-Sinai Medical Center
: TITLE OF INVENTION: The regents of the University of California
: TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
: FILE REFERENCE: PP-PM 4713
: CURRENT APPLICATION NUMBER: PCT/US01/16032
: CURRENT FILING DATE: 2001-05-17
: PRIOR APPLICATION NUMBER: US 09/575,061
: PRIOR FILING DATE: 2000-05-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 100
: TYPE: PR
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Microbial organism from the human gut
PCT-US01-16032-3

```

Query Match	100.0%	Score 493;	DB 1;	length 100;
Best Local Similarity	100.0%	Pred. No. 4.6e-51;		
Matches 100; Conservative	0;	Mismatches	0;	Gaps 0

**Oy** 1 DLASAVGIQSSSIFHHHRSKDEILRAVMETIHYNTAMMRASLEASTYRERVALLIRCE 600  
**Dd** 1 DLASAVGIQSSSIFHHHRSKDEILRAVMEETIHYNTAMMRASLEASTYRERVALLIRCE 600

Qy	Db
61 LOSIMGSSGEAMAVLVYEMRSLAEGQAHVLAIRDVYEI 1000	61 LOSIMGSGEAMAVLVYEMRSLAEGQAHVLAIRDVYEI 1000

## 3

```

PCT-US01-16032A-3
: Sequence 3, Application PC/TUS0116032A
:
: GENERAL INFORMATION:
:
: APPLICANT: Targan, Stephen R.
:
: APPLICANT: Braun, Jonathan
:
: APPLICANT: Sutton, Christopher L.
:
: TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
:
: TITLE OF INVENTION: Crohn's Disease Using The Ompc Antigen
:
: FILE REFERENCE: FP-PW 4713
:
: CURRENT APPLICATION NUMBER: PCT/US01/16032A
:
: CURRENT FILING DATE: 2001-05-17
:
: PRIOR APPLICATION NUMBER: US 09/575,061
:
: PRIOR FILING DATE: 2000-05-19
:
: NUMBER OF SEQ ID NOS: 3
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO: 3
:
: LENGTH: 100
:
: TYPE: PRT
:
: ORGANISM: Unknown
:
: FEATURE:
:
: OTHER INFORMATION: Microbial organism from the human gut
:
PCT-US01-16032A-3

```

Query Match	100.0%	Score 493;	DB 1;	Length 100;
Best Local Similarity	100.0%	Pred. No. 4.6e-51;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	DLASAVGIQGGSIFFHFRFSKDEILEAVMEETHTHTNTAMRASLEEVSTYREKVALIRCE	60	

Dd 1 DLSANGIQSGSTFHHFKKDELRLAWEDETHINTAMMRASLEASTVERVALLICE 600

Qy 61 LQSI MGSGEAMAVLYEKRSLSAEGQAHVALRDVEQI 100  
|||||  
61 LQSI MGSGEAMAVLYEKRSLSAEGQAHVALRDVEQI 100

Dd 61 LQSI MGSGEAMAVLYEKRSLSAEGQAHVALRDVEQI 100

## 4

```

US-09-303-120B-2
: Sequence 2, Application US/09303120B
: GENERAL INFORMATION:
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher
: TITLE OF INVENTION: 1BD-Associated Microbial Antigens and Methods of Using
: TITLE OF INVENTION: Same
: FILE REFERENCE: P-PM 3478
: CURRENT APPLICATION NUMBER: US/09/303,120B
: CURRENT FILING DATE: 1999-04-30
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 100
: TYPE: PR1
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: Microbial
: US-09-303-120B-2

```

Query Match	100.0%	Score 493;	DB 17	Length 100;
Best Local Similarity	100.0%	Pred. No. 4.6e-51;		
Best Match	100;	Conservative 0;	Mismatches 0;	Indels 0;

```
OY      1 DLASAVGIQSSIFHHFKSKDELIRAVEETIHYNTAMMRASLEASTVREVLALRCE    600
|||||
|||||
|||||
Db      1 DLASAVGIQSSIFHHFKSKDELIRAVEETIHYNTAMMRASLEASTVREVLALRCE    600
```

Qy 61 LQSIMGSGEANAALVYEWRSLSAEGGAHVLAIRDVYEDI 1000  
|||||  
Db 61 LQSIMGSGEANAALVYEWRSLSAEGGAHVLAIRDVYEDI 1000

५

```

RESULT      5
US-09-575-061-3
: Sequence 3: Application US/09575061
: GENERAL INFORMATION:
: APPLICANT: Targan, Stephan R.
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher L.
: TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
: TITLE OF INVENTION: Crohn's Disease Using The Ompc Antigen
: FILE REFERENCE: P-PM 4097
: CURRENT APPLICATION NUMBER: US/09/575,061
: CURRENT FILING DATE: 2000-05-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 100
: TYPE: PR1
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Microbial organism from the human gut
US-09-575-061-3

```

Query Match	100.0%	Score 493;	DB 19;	Length 100;
Best Local Similarity	100.0%	Pred. No. 4,6e-51;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Caps 0;
Oy	1	DLASAVGIQSSIFPHFESKDEILRAVNEETHYNTAMMRASLEASTYREYVALLICE	60	
1	DLASVGIQSSIFPHFESKDEILRAVNEETHYNTAMMRASLEASTYREYVALLICE	60		

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100  
DB 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

RESULT 6  
US-09-820-576-2  
; Sequence 2, Application US/09820576  
; GENERAL INFORMATION:  
; APPLICANT: Braun, Jonathan  
; TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid  
; FILE REFERENCE: P-PM 4646  
; CURRENT APPLICATION NUMBER: US/09/820,576  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 09/303,120  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Microbial organism from the human gut  
US-09-820-576-2

Query Match 100.0%; Score 493; DB 22; Length 100;  
Best Local Similarity 100.0%; Pred. No. 4,66-51;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60  
DB 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100  
DB 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

RESULT 7  
US-09-252-991A-26292  
; Sequence 26292, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196,136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26292  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26292

Query Match 87.0%; Score 429; DB 16; Length 245;  
Best Local Similarity 84.0%; Pred. No. 8,2e-43;  
Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60  
DB 87 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 146  
QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

DB 147 LOSINGSGEAMAVLYEWRSLAEGQAVIILGLRIYEQM 186

RESULT 8  
US-09-328-352-5572  
; Sequence 5572, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; PRIOR FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5572  
; LENGTH: 206  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5572

Query Match 54.2%; Score 267; DB 17; Length 206;  
Best Local Similarity 54.0%; Pred. No. 2,2e-23;  
Matches 54; Conservative 21; Mismatches 25; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60  
DB 50 ELAOFIIGIOSGSLFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 109  
QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100  
DB 110 LISTIGDTGAMAVLYEWRSLAEGQAVLALRDVEEI 149

RESULT 9  
PCT-US00-11473-6  
; Sequence 6, Application PC/TUS0011473  
; GENERAL INFORMATION:  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: IBD-Associated Microbial Antigens and Methods of Using  
; FILE REFERENCE: P-PM 4142  
; CURRENT APPLICATION NUMBER: PCT/US00/11473  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 09/303,120  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 190  
; TYPE: PRT  
; ORGANISM: Clostridium pasteurianum  
PCT-US00-11473-6

Query Match 22.4%; Score 110.5; DB 1; Length 190;  
Best Local Similarity 27.0%; Pred. No. 0.00014;  
Matches 27; Conservative 25; Mismatches 35; Indels 13; Gaps 3;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60  
DB 28 ELASNAAGVAKGTLTYHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 83  
QY 57 INCELOSINGSGEAMAVLYEWRSLAEGQAVLALRDV 96  
DB 84 CRVQNLNLYKNRDFKVIASQIMGR-----ELRQLELDI 118

RESULT 10  
US-09-303-120B-6  
; Sequence 6, Application US/09303120B  
; GENERAL INFORMATION:

```

: RESULT 12
: US-09-902-540-14870
: Sequence 14870, Application US/09902540
: GENERAL INFORMATION:
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

```

```

RESULT 14
US-09-303-120B-7
Sequence 7, Application US/09303120B
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Sutton, Christopher
TITLE OF INVENTION: IBID-Associated Microbial Antigens and Methods of Using
FILE REFERENCE: P-PM 3478
CURRENT APPLICATION NUMBER: US/09/303,120B
CURRENT FILING DATE: 1999-04-30
NUMBER OF SEQ. ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 7
;
; LENGTH: 200
;
; TYPE: PRF
;
; ORGANISM: Mycobacterium tuberculosis
US-09-303-1208-7

```

Query Match 19.2%; Score 94.5; DB 17; Length 200;  
Best Local Similarity 26.9%; Pred. No. 0.013.  
Matches 21; Conservative 20; Mismatches 34; Indels 3; Gaps 1;

Qy 1 DLASVGIQSSIFHHFSKDELRAVMEETIHYNAMMRASLEEAASYRERVALLIRCE 60  
 1:1 11 111:11 11:: : : : 1 : : : 11: 1  
 Db 35 DIADGAGILSGSLYHHFASKKEWDELKGLDWLFARYRDIYDSTANPDLERLOGIFMAS 94

```
Oy      61 LQIMGSGEAMVLYE 78
          ::|      :|::|:
Db      95 FEAEIHHAQ---VVIYQ 109
```

RESULT 15  
US-09-820-

```

: Sequence 7, Application US/09820576
: GENERAL INFORMATION:
: APPLICANT: Braun, Jonathan
: APPLICANT: Brauton, Christopher L.
: TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid
: TITLE OF INVENTION: Molecules
: FILE REFERENCE: P-PM 4646
: CURRENT APPLICATION NUMBER: US/09/820,576
: CURRENT FILING DATE: 2001-03-28
: PRIOR APPLICATION NUMBER: US 09/303,120
: PRIOR FILING DATE: 1999-04-30
: NUMBER OF SEQ. ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 200
: TYPE: PRY
: ORGANISM: Mycobacterium tuberculosis
: US-09-820-576--7

```

Query Match	19.2%;	Score 94.5;	DB 22;	Length 200;
Best Local Similarity	26.9%;	Pred. No. 0.013;		
Matches 21; Conservative	20;	Mismatches 34;	Indels 3;	Gaps 1;

**OY**    1 DLASAVGIQSSIFHHFRSKDELRAVMEETIHYNAMRASLEASTYRERVALLIRCE    60  
      | : | | | | | | : : : : | : : |  
**DG**    35 DIADAGAGILSGSLYHFFASKKEAWDELLRGFLDWLPARYRDIVDSTANPLERLGLFMAS    94

```

OY      61 LQSIMGSGEAMAVLYE 78
        ::! : '!::! :
Db      95 FEAEIEHHAQ--VVITYQ 109

```

Search completed: March 4, 2002, 20:27:16  
Job time: 184 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:24:37 ; Search time 19.94 Seconds  
(without alignments)  
287.207 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIGSGSIFHFKSK.....SLSAEGAHVLAIRDVEEQI 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 261336 seqs, 57268997 residues

Total number of hits satisfying chosen parameters: 261336

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCIT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	100.0	100	5	US-09-966-608-2
2	484	98.2	198	5	US-09-976-451-5
3	110.5	22.4	190	5	US-09-966-608-6
4	94.5	19.2	200	5	US-09-966-608-7
5	92.5	18.8	193	5	US-09-897-516-4731
6	90	18.3	218	5	US-09-605-703B-2830
7	89	18.1	192	5	US-09-966-608-8
8	89	18.1	576775	4	US-08-895-611-2
9	89	18.1	576775	4	US-08-895-611D-2
10	89	18.1	576775	5	US-09-895-611D-2
11	86	17.4	189	6	US-10-015-127-12199
12	84	17.0	196	5	US-09-897-516-4814
13	79.5	16.1	183	5	US-09-897-516-5762
14	79.5	16.1	193	5	US-09-897-516-5915
15	77.5	15.7	194	5	US-09-815-242-5102
16	75	15.2	1489	5	US-09-614-150-6636
17	69	14.0	80	5	US-09-545-199D-129
18	68.5	13.9	221	5	US-09-708-427-8313
19	68.5	13.9	391	5	US-09-708-427-8313
20	68	13.8	205	6	US-10-015-127-11169
21	68	13.8	262	5	US-09-708-427-52027
22	68	13.8	320	5	US-09-708-427-52026
23	68	13.8	509	6	US-10-018-366-5
24	67.5	13.7	186	6	US-10-015-127-12405
25	67.5	13.7	199	5	US-09-815-242-13852
26	67	13.6	131	6	US-10-015-127-11523

27	67	13.6	187	5	US-09-897-516-6487	Sequence 6487, Ap
28	66.5	13.5	199	5	US-09-815-242-10169	Sequence 10169, A
29	66.5	13.5	349	5	US-09-675-784A-13026	Sequence 13026, A
30	66	13.4	182	6	US-10-015-127-11354	Sequence 11354, A
31	66	13.4	219	5	US-09-897-516-7388	Sequence 7388, Ap
32	65.5	13.3	195	5	US-09-151-957-2	Sequence 2, Appl1
33	65.5	13.3	384	5	US-09-708-427-14867	Sequence 14867, A
34	65.5	13.3	419	5	US-09-708-427-14866	Sequence 14866, A
35	65.5	13.3	425	5	US-09-708-427-14865	Sequence 14865, A
36	65	13.2	957	5	US-09-897-516-7494	Sequence 7494, Ap
37	64.5	13.1	215	7	US-60-343-278-69	Sequence 69, Appl
38	64	13.0	210	6	US-10-015-127-11511	Sequence 11511, A
39	64	13.0	233	5	US-09-815-242-5121	Sequence 5121, Ap
40	64	13.0	617	5	US-09-604-693A-64	Sequence 64, Appl
41	63.5	12.9	912	7	US-60-337-358-481	Sequence 481, Appl
42	63	12.8	350	5	US-09-800-198-60	Sequence 60, Appl
43	62.5	12.7	1380	5	US-09-614-150-1101	Sequence 1101, Ap
44	62	12.6	97	6	US-10-015-127-9935	Sequence 9935, Ap
45	62	12.6	219	5	US-09-921-650B-32	Sequence 32, Appl

## ALIGNMENTS

```

RESULT 1
US-09-966-608-2
; Sequence 2, Application US/09966608
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid
; FILE REFERENCE: P-PM 4966
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 09/303,120
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 09/820,576
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Microbial organism from the human gut
US-09-966-608-2

Query Match      100.0%  Score 493:  DB 5:  Length 100:
Best Local Similarity 100.0%:  Pred. No. 3.1e-48:
Matches 100:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

Oy 1 DLASAVGIGSGSIFHFKSKDELIRAVMEETIHYNTAMKRSLEASVYREVLALIRCE 60
    |||||||
Db 1 DLASAVGIGSGSIFHFKSKDELIRAVMEETIHYNTAMKRSLEASVYREVLALIRCE 60
    |||||||
Oy 61 LOSIMGSGEAMAVLYEWRSLSAEGAHVLAIRDVEEQI 100
    |||||||
Db 61 LOSIMGSGEAMAVLYEWRSLSAEGAHVLAIRDVEEQI 100
    |||||||

RESULT 2
US-09-976-451-5
; Sequence 5, Application US/09976451
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Mei, Bo
; APPLICANT: Forbes, Ashley
; TITLE OF INVENTION: Methods of Diagnosing and Treating
; TITLE OF INVENTION: Crohn's Disease using Pseudomonas Antigens
; FILE REFERENCE: P-PM 4968

```

05-10-20  
BCH

Query Match	19.2%	Score 94.5;	DB 5;	Length 200;
Best Local Similarity	26.9%	Pred. NO. 0.0045;		

[illegible]

```

RESULT      5
US-09-897-516-4731
; Sequence 4731, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spolidonov, Sergei
; TITLE OF INVENTION: Xenothadus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4731
; LENGTH: 193

```

```

US-09-897-516-4731

Query Match      18.8%  Score 92.5; DB 5; Length 193;
Best Local Similarity 30.6%  Pred. No. 0.0071;
Matches 33; Conservative 25; Mismatches 39; Indels 11; Gaps 6;

OY 1 DLASAVGIQSSIFTHHFKSKDELRAVEETIHTYNTAMKRSLEAS--TYREVALI 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 ELVAAGAPKCSFYFFKSKKEFGQILKE--YFENEYLSRAAMLLSTEGTAREHLN 88

```

```

Db      89  RVCSEIQTGQPEDKCLTVKLGAEVCDDL5-BGMRSILLCRGTVDIIDQL 135

RESULT      6
US-09-605-703B-2830
: Sequence 2830, Application US/09605703B
: GENERAL INFORMATION:
: APPLICANT: Pompejus, Markus
: APPLICANT: Kroger, Burkhard
: APPLICANT: Schroder, Hartwig

```



```

RESULT 10
US-09-895-611D-2
: Sequence 2, Application US/09895611D
:
: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: Swanson, Ronald V.
: APPLICANT: Short, Jay M.
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF THE AQUIFEX
: TITLE OF INVENTION: ACOLICUS GENOME, FRAGMENTS THEREOF, AND USES THEREOF
: FILE REFERENCE: DIVER1330
: CURRENT APPLICATION NUMBER: US/09/895,611D
: CURRENT FILING DATE: 1997-07-16
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 576775
:
: TYPE: PRF
: ORGANISM: Aquifex aeolicus
US-09-895-611D-2

```



OY 60 ELOSINGSGEAMAVLYEWMRLSAEGOAHVLAIRDVEO 99  
DB 95 NFDTSQINS-AAMKTWIAFW-ASSMHQPSLYRLQOVNER 131

RESULT 15  
US-09-815-242-5102

; Sequence 5102, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5102  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5102

Query Match 15.7%; Score 77.5; DB 5; Length 194;  
Best Local Similarity 25.0%; Pred. No. 0.35;  
Matches 25; Conservative 23; Mismatches 43; Indels 9; Gaps 3;

OY 1 DIASAVGIQSGSIFHFHFKSKDELIRAVMEETH-INTAMRASLEASTVREPVIALIRC 59  
DB 32 EILQSGAGVPKGSFYHFFKSKQEGQALLDFFRVYLAIDMDORFSAPGLNARERLMSYQK 91  
OY 60 ELOSINGSGEAMAVLYEWMRLSAE---GOAHVLAIRD 95  
DB 92 WIDNACPPDEORCLV---KLSAEVADLSRMTLRD 127

Search completed: March 4, 2002, 20:31:16  
Job time: 399 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:22:32 ; Search time 27.71 Seconds  
(without alignments)  
267.316 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIGSGSIFHHFKSK.....SLSAEGCAHYALRDVEQL 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_1101.\*

- 1: /SID2/gcgdata/geneseq/genesep/AA1980.DAT.\*
- 2: /SID2/gcgdata/geneseq/genesep/AA1981.DAT.\*
- 3: /SID2/gcgdata/geneseq/genesep/AA1982.DAT.\*
- 4: /SID2/gcgdata/geneseq/genesep/AA1983.DAT.\*
- 5: /SID2/gcgdata/geneseq/genesep/AA1984.DAT.\*
- 6: /SID2/gcgdata/geneseq/genesep/AA1985.DAT.\*
- 7: /SID2/gcgdata/geneseq/genesep/AA1986.DAT.\*
- 8: /SID2/gcgdata/geneseq/genesep/AA1987.DAT.\*
- 9: /SID2/gcgdata/geneseq/genesep/AA1988.DAT.\*
- 10: /SID2/gcgdata/geneseq/genesep/AA1989.DAT.\*
- 11: /SID2/gcgdata/geneseq/genesep/AA1990.DAT.\*
- 12: /SID2/gcgdata/geneseq/genesep/AA1991.DAT.\*
- 13: /SID2/gcgdata/geneseq/genesep/AA1992.DAT.\*
- 14: /SID2/gcgdata/geneseq/genesep/AA1993.DAT.\*
- 15: /SID2/gcgdata/geneseq/genesep/AA1994.DAT.\*
- 16: /SID2/gcgdata/geneseq/genesep/AA1995.DAT.\*
- 17: /SID2/gcgdata/geneseq/genesep/AA1996.DAT.\*
- 18: /SID2/gcgdata/geneseq/genesep/AA1997.DAT.\*
- 19: /SID2/gcgdata/geneseq/genesep/AA1998.DAT.\*
- 20: /SID2/gcgdata/geneseq/genesep/AA1999.DAT.\*
- 21: /SID2/gcgdata/geneseq/genesep/AA2000.DAT.\*
- 22: /SID2/gcgdata/geneseq/genesep/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	100.0	100	21	AA19839
2	90	18.3	222	22	AAG90710
3	83	16.8	214	22	AAU04043
4	78.5	15.9	118	21	AA19839
5	73	14.8	3722	12	AA19839
6	70.5	14.3	188	22	AA19839
7	69	14.0	80	21	AA19839
8	69	14.0	2818	13	AA19839
9	68	13.8	509	19	AA19839
10	68	13.8	509	19	AA19839
11	68	13.8	509	21	AA19839

12	67.5	13.7	508	17	AA19839	Protein having pro
13	67.5	13.7	508	18	AA19839	Arabidopsis thalia
14	67.5	13.7	508	18	AA19839	Arabidopsis thalia
15	67.5	13.7	508	19	AA19839	Arabidopsis thalia
16	67.5	13.7	508	19	AA19839	Arabidopsis thalia
17	67.5	13.7	508	22	AA19839	Arabidopsis thalia
18	66.5	13.5	136	22	AA19839	Arabidopsis thalia
19	66.5	13.5	195	19	AA19839	Arabidopsis thalia
20	65.5	13.3	411	19	AA19839	Arabidopsis thalia
21	65.5	13.3	414	19	AA19839	Arabidopsis thalia
22	65.5	13.3	444	19	AA19839	Arabidopsis thalia
23	64.5	13.1	195	21	AA19839	Arabidopsis thalia
24	64.5	13.1	326	21	AA19839	Arabidopsis thalia
25	64	13.0	156	21	AA19839	Arabidopsis thalia
26	64	13.0	783	22	AA19839	Arabidopsis thalia
27	64	13.0	3768	12	AA19839	Arabidopsis thalia
28	64	13.0	3778	12	AA19839	Arabidopsis thalia
29	63.5	12.9	1784	16	AA19839	Arabidopsis thalia
30	63	12.8	2485	15	AA19839	Arabidopsis thalia
31	63	12.8	2485	15	AA19839	Arabidopsis thalia
32	63	12.8	2818	18	AA19839	Arabidopsis thalia
33	63	12.8	2818	22	AA19839	Arabidopsis thalia
34	62.5	12.7	82	21	AA19839	Arabidopsis thalia
35	62.5	12.7	224	21	AA19839	Arabidopsis thalia
36	62.5	12.7	76	22	AA19839	Arabidopsis thalia
37	62	12.6	217	22	AA19839	Arabidopsis thalia
38	62	12.6	219	22	AA19839	Arabidopsis thalia
39	62	12.6	435	22	AA19839	Arabidopsis thalia
40	62	12.6	483	22	AA19839	Arabidopsis thalia
41	62	12.6	778	21	AA19839	Arabidopsis thalia
42	62	12.6	806	21	AA19839	Arabidopsis thalia
43	62	12.6	927	21	AA19839	Arabidopsis thalia
44	61.5	12.5	76	14	AA19839	Arabidopsis thalia
45	61.5	12.5	211	22	AA19839	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	AA19839	standard; Protein: 100 AA.
ID	AA19839	
XX	AA19839	
AC	AA19839	
XX	AA19839	
DT	05-MAR-2001	(first entry)
XX	05-MAR-2001	
DE	Inflammatory bowel disease associated antigen I-2.	
XX	Inflammatory bowel disease associated antigen I-2.	
KW	I-2; Inflammatory bowel disease; IBD; Crohn's disease; therapy;	
XX	diagnosis; vaccine.	
OS	Unidentified microorganism.	
XX	Unidentified microorganism.	
XX	WO200066067-A2.	
PN	WO200066067-A2.	
XX	09-NOV-2000.	
PD	09-NOV-2000.	
XX	28-APR-2000; 2000MO-US11473.	
PF	28-APR-2000; 2000MO-US11473.	
XX	30-APR-1999; 99US-0303120.	
PR	30-APR-1999; 99US-0303120.	
XX	(RECC ) UNTV CALIFORNIA.	
PA	(RECC ) UNTV CALIFORNIA.	
XX	Braun J, Sutton C;	
PI	Braun J, Sutton C;	
XX	WPI: 2000-687440/67.	
DR	WPI: 2000-687440/67.	
XX	N-PSDB: AAA89029.	
XX	N-PSDB: AAA89029.	
PT	Inflammatory bowel disease (IBD) associated I-2 polypeptides useful for	
PT	diagnosing IBD, vaccinating against IBD and for identifying agents for	
PT	treating IBD -	
XX	treating IBD -	

PS Claim 1; Fig 1A; 76pp; English.  
XX  
CC The present sequence is that of novel microbial I-2 polypeptide.  
CC Representative difference analysis was used to isolate DNA  
CC sequences from a Crohn's disease (CD) patient that were  
CC differentially present in mononuclear cells from the lamina propria  
CC in an area with ulcerations as compared to an area free of disease.  
CC 2 inflammatory bowel disease (IBD)-associated sequences were of  
CC microbial origin, designated I-1 (see AAB19840-41) and I-2. I-2  
CC sequences were more often found in involved CD tissue than in  
CC ulcerative colitis or non-IBD samples. The I-2 polypeptide may be  
CC used in methods of the invention for treating IBD, diagnosing IBD,  
CC vaccinating against IBD, and identifying candidate agents useful  
CC for treating IBD, especially CD.  
SQ  
XX  
SQ Sequence 100 AA;  
Query Match 100.0%; Score 493; DB 21; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.4e-54;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLSAVGIQSGSIFHFHFKSKDEILRAVMEETIHYNTAMMRASLEASVRRVLAIRCE 60  
Db 1 dlsavgiqsgsifhfhfkskdeilravmeethyntammrassleasvrrvlla1rce 60  
QY 61 LOSINGSGSEAMAVLYEWRSLAEGQAVLALRDVYEQI 100  
Db 61 lqslngsgseamavlyewrslasegqavhaldvdyeqi 100  
RESULT 2  
AAG90710  
ID AAG90710 standard; Protein: 222 AA.  
XX  
AC AAG90710;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamylcmm protein fragment SEQ ID NO: 4464.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EPI108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
DR WPI: 2001-376931/40.  
DR N-PSDB: AAH65929.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
PS Claim 17; SEQ ID NO: 4464; 246pp + sequence listing; English.  
CC  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
SQ  
XX  
SQ Sequence 222 AA;  
Query Match 18.3%; Score 90; DB 22; Length 222;  
Best Local Similarity 29.4%; Pred. No. 0.0028;  
Matches 20; Conservative 20; Mismatches 20; Indels 8; Gaps 2;  
QY 2 LASAVGIQSGSIFHFHFKSKDEILRAVMEETIHYNTAMMRASLEASV---RRVLAIR 57  
Db 45 ladvglrqaelyhfhfkskdeilrlktsvepstvl----aedlstldgpenrlwaly 100  
QY 58 RCEIQSTIM 65  
Db 101 asevrlll 108  
RESULT 3  
AAU04043  
ID AAU04043 standard; Protein: 214 AA.  
XX  
AC AAU04043;  
XX  
DT 23-OCT-2001 (first entry)  
XX  
DE Streptococcus coelicolor Mmfr protein.  
XX  
KW SCP1; methylenomycin cluster; mmc; MmyR; Mmfp; Mmfi;  
KW Mmfl; Mmfr; MmyT; Mmyo; Mmyg; MmyJ; Mmr; heterologous gene expression.  
XX  
OS Streptococcus coelicolor.  
XX  
PN WO200148228-A1.  
XX  
PD 05-JUL-2001.  
XX  
PF 20-DEC-2000; 2000WO-GB04972.  
XX  
PR 23-DEC-1999; 99GB-0030477.  
XX  
PA (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
PI Chater KF, Bruton CJ, O'Rourke SJ, Wietzorrek AW;  
XX  
XX WPI: 2001-425675/45.  
DR N-PSDB: AAS07627.  
DR AAU04045, AAU04046.  
XX  
PT Novel expression cassette for expressing a nucleic acid of interest,  
PT derived from the regulatory region of methylenomycin gene cluster of  
PT SCP1 plasmid of Streptomyces coelicolor A3(2)  
XX  
PS Claim 25; Fig 8e; 142pp; English.  
XX  
CC The sequence represents the Mmfr protein encoded by the mmfr gene carried  
CC on the expression cassette present on plasmid SCP1. The expression  
CC cassette is the regulatory region of the methylenomycin cluster (mmc)  
CC from Streptomyces coelicolor A3(2), which encodes the MmyR, Mmfp, Mmfi,  
CC Mmfl, Mmfr, MmyT, Mmyo, Mmyg, MmyJ and partial Mmr polypeptides. The  
CC expression cassette is useful for expressing a nucleic acid of interest,  
CC substantially only when the host cell culture reaches high cell density  
CC at or close to the stationary phase of host cell culture. In particular

AAC/4446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,

PT Prepn. of cephalosporin series antibiotics - comprises culturing  
PT transformant of microbe transformed by plasmid contg. new DNA  
PT fragment







[illegible][illegible]

```

PT DNA containing a plant proto-porphyrinogen oxidase gene promoter
PT optionally linked to a heterologous gene, especially to express
PT herbicide-resistant enzymes, and plants containing such constructs
XX
PS Claim 27: Pages 43-46; 114pp; English.
XX
CC The present sequence is Arabidopsis thaliana
CC protoporphyrinogen oxidase-2 (protox-2).
CC The protox-1 promoter can be used to express herbicide resistant
CC enzymes, specifically protox, i.e. a plant tissue, plant or progeny
CC containing a chimeric gene of the promoter and a heterologous
CC coding sequence. The plant can also be used in breeding programmes.
CC Also hybridising fragments of the protox coding sequence can be
CC used as probes, e.g. to isolate related genes or for genomic
CC mapping.
XX
XQ Sequence 508 AA:
XX
Query Match 13.7% Score 67.5 DB 18 Length 508:
Best Local Similarity 26.7% Pred No 6:
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps
QY 15 HHRKSDLEILRAVM-----ETIHYNTAMMRAS---LEEASTVERVALIRCELOSIM 65
| | | : : : : | : : | : : : : : : : : : : : : : : : : : : :
Db 366 hgffltlflssmmfdpsdvhltytftlgsgrngelakast--delkqvtsldqrll 422
QY 66 GGSGEAMAVLYVEWR 80
| | | : : | | |
Db 424 gvegepvsvnyhywr 438
.
RESULT 14
AAW25747
ID AAW25747 standard; Protein: 508 AA.
AC AAW25747;
AD
DT 01-MAR-1998 (first entry)
DE Arabidopsis protoporphyrinogen oxidase (protox-2).
DS
KW Protox-2; protoporphyrinogen oxidase; inhibitor;
KM herbicide tolerance; herbicide resistance; transgenic plant.
OS Arabidopsis thaliana.
XX
XX MO9732011-A1.
XX
PD 04-SEP-1997.
XX
PF 27-FEB-1997; 97WO-US03313.
XX
RR 21-JUN-1996; 96US-0020003.
RR 28-FEB-1996; 96US-0012705.
RR 28-FEB-1996; 96US-0013612.
PA (NOVS ) NOVARTIS AG.
PI Helfetz PB, Johnson MA, Potter SL, Volrath SL, Ward ER;
DR WPI: 1997-AA8683/41.
DR N-PSDB; AAT86130.
XX
PT New DNA encoding plant protoporphyrinogen oxidase enzyme - and
PT herbicide resistant mutants, useful to prepare plants resistant to
PT herbicides which therefore kills undesired vegetation only
XX
PS Claim 95: Page 102-105; 196pp; English.
XX

```

CC was deduced from a previously obtained Protox-1 cDNA clone (see  
 CC AAT86130). Arabidopsis protox-2 can be modified to render it  
 CC resistant to protox inhibitors and hence tolerant of certain  
 CC herbicides. Plants, especially crop plants, may be engineered for  
 CC resistance to protox inhibitors via mutation of the native protox  
 CC gene to a resistant form, or they may be transformed with a gene  
 CC encoding an inhibitor-resistant form of a plant protox enzyme, such  
 CC as claimed forms from wheat, soybean, cotton, sugarbeet, oilseed  
 CC rape, rice and sorghum (see AAW25738-48). Application of herbicide  
 CC will then kill undesired vegetation only. Protox enzymes can also  
 CC be expressed in transformed host cells and used to identify  
 CC inhibitors of protox enzyme activity, i.e. herbicide candidates, or  
 CC to design herbicide tolerant forms of the enzyme.

xx Sequence 508 AA:

Query Match 13.7%; Score 67.5; DB 18; Length 508;

Best Local Similarity 26.7%; Pred. No. 6;  
 Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

OY 15 HHFKSKDELRAVM-----ETIHYNTAMMRAS---LEASTVNERVLAIRCLOSIM 65  
 | | | : : : | : | : | : | : | : : : : | : :  
 Db 366 hgfkclglfssmmfpdrpsdvhltyltfigsrngelakast--delkqvtsdlqrll 423

OY 66 GSGGEAMAVLYEWR 80  
 | | : : | | | |  
 Db 424 gvegepvsvnhyywr 438

RESULT 15

AAWS1254  
 ID AAW51254 standard; Protein; 508 AA.

AC AAW51254;

DT 14-AUG-1998 (first entry)

DE Arabidopsis proto-porphyrinogen oxidase-2 (protox-2).

KW Proto-porphyrinogen oxidase; protox; herbicide tolerant; resistance;

XX Inhibitor; variegate porphyrin.

OS Arabidopsis thaliana.

XX US5767373-A.

PD 16-JUN-1998.

PF 16-JUN-1994; 94US-0261198.

PR 06-JUN-1995; 95US-0472028.

PR 16-JUN-1994; 94US-0261198.

XX (NOVS ) NOVARTIS FINANCE CORP.

PI Volrach S, Ward ER;

DR WPI; 1998-361821/31.

DR N-PSDB; AAV07252.

PT DNA encoding eukaryote herbicide resistant proto-porphyrinogen  
 PT oxidase - useful for producing recombinant plants having functional  
 PT enzyme, to be grown in the presence of herbicides

PS Claim 6; Columns 21-24; 43pp; English.

CC The invention relates to eukaryotic DNA sequences coding for native  
 CC proto-porphyrinogen oxidase (protox) or modified forms of the enzyme  
 CC which are herbicide tolerant. Plants having altered protox activity  
 CC which confers tolerance to herbicides are also provided. These plants  
 CC may be bred or engineered for resistance to protox inhibitors via  
 CC mutation of the native protox gene to a resistant form or through

CC increased levels of expression of the native protox gene, or they may be  
 CC transformed with modified eukaryotic or prokaryotic protox coding  
 CC sequences or wild type prokaryotic protox sequences which are herbicide  
 CC tolerant. In the human condition variegate porphyrin, an autosomal  
 CC dominant disorder characterized by neuropsychiatric symptoms, decreased  
 CC protox activity is detected. Thus the DNA molecules can be used to  
 CC provide probes to detect and quantify protox levels in the diagnosis of  
 CC diseases associated with the enzyme. The present sequence represents  
 CC Arabidopsis proto-porphyrinogen oxidase-2 (protox-2).

xx Sequence 508 AA:

Query Match 13.7%; Score 67.5; DB 19; Length 508;

Best Local Similarity 26.7%; Pred. No. 6;  
 Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

OY 15 HHFKSKDELRAVM-----ETIHYNTAMMRAS---LEASTVNERVLAIRCLOSIM 65  
 | | | : : : | : | : | : | : | : : : : | : :  
 Db 366 hgfkclglfssmmfpdrpsdvhltyltfigsrngelakast--delkqvtsdlqrll 423

OY 66 GSGGEAMAVLYEWR 80  
 | | : : | | | |  
 Db 424 gvegepvsvnhyywr 438

Search completed: March 4, 2002, 20:24:06  
 Job time: 94 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:23:12 ; Search time 15.26 Seconds  
(without alignments)  
147.466 Million cell updates/sec

Title: US-09-966-608-2  
Perfect score: 493  
Sequence: 1 DLASAVGIGSGIFHHFKSK.....SLSAEQAHVLRDVEQI 100

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	13.8	503	4	US-08-740-223A-11 Sequence 11, Appl
2	68	13.8	509	2	US-08-665-926-8 Sequence 8, Appl
3	68	13.8	509	4	US-08-740-223A-10 Sequence 10, Appl
4	67.5	13.7	508	1	US-08-472-028A-4 Sequence 4, Appl
5	67.5	13.7	508	2	US-08-808-931-4 Sequence 4, Appl
6	67.5	13.7	508	3	US-08-808-323-4 Sequence 4, Appl
7	67.5	13.7	508	3	US-09-050-603A-4 Sequence 4, Appl
8	67.5	13.7	508	3	US-09-102-420B-4 Sequence 4, Appl
9	67.5	13.7	508	4	US-09-071-296-4 Sequence 4, Appl
10	67.5	13.7	508	4	US-09-196-268-4 Sequence 4, Appl
11	67.5	13.7	508	4	US-09-015-683-4 Sequence 4, Appl
12	67.5	13.7	508	4	US-09-191-998-4 Sequence 4, Appl
13	67	13.6	2818	2	US-08-449-933-2 Sequence 2, Appl
14	67	13.6	2818	4	US-07-966-049A-2 Sequence 2, Appl
15	65.5	13.3	195	2	US-08-882-704A-2 Sequence 2, Appl
16	64	13.0	768	2	US-08-222-617A-5 Sequence 5, Appl
17	64	13.0	3666	2	US-08-222-617A-12 Sequence 12, Appl
18	64	13.0	3727	2	US-08-222-617A-27 Sequence 27, Appl
19	64	13.0	3778	2	US-08-222-617A-2 Sequence 2, Appl
20	63.5	12.9	1784	4	US-09-040-738-2 Sequence 2, Appl
21	63.5	12.9	1784	4	US-08-652-426A-2 Sequence 2, Appl
22	63	12.8	2485	5	PCT-US94-00198-1 Sequence 1, Appl
23	63	12.8	2485	5	PCT-US94-00198-2 Sequence 2, Appl
24	63	12.8	2818	1	US-08-510-284-1 Sequence 1, Appl
25	63	12.8	2818	1	US-08-411-389-2 Sequence 2, Appl
26	63	12.8	2818	1	US-09-542-331-2 Sequence 2, Appl
27	62	12.6	219	4	US-09-162-184-32 Sequence 32, Appl

28	62	12.6	219	4	US-09-489-777A-32 Sequence 32, Appl
29	61.5	12.5	76	2	US-08-244-537-6 Sequence 6, Appl
30	61.5	12.5	2325	3	US-08-417-089-6 Sequence 6, Appl
31	61.5	12.5	2325	4	US-08-695-651-6 Sequence 6, Appl
32	61.5	12.5	2325	4	US-08-930-285-6 Sequence 6, Appl
33	61.5	12.5	2325	4	US-08-695-421-6 Sequence 6, Appl
34	59.5	12.1	216	4	US-09-162-184-31 Sequence 31, Appl
35	59.5	12.1	216	4	US-09-489-777A-31 Sequence 31, Appl
36	59	12.0	218	4	US-09-162-184-30 Sequence 30, Appl
37	59	12.0	218	4	US-09-489-777A-30 Sequence 30, Appl
38	58.5	11.9	785	1	US-07-841-997A-4 Sequence 4, Appl
39	58.5	11.9	785	1	US-08-290-301-4 Sequence 4, Appl
40	58.5	11.9	1098	1	US-08-290-301-82 Sequence 82, Appl
41	58	11.8	295	1	US-08-317-223-1 Sequence 1, Appl
42	58	11.8	295	3	US-09-059-849A-1 Sequence 1, Appl
43	58	11.8	295	3	US-09-213-632-1 Sequence 1, Appl
44	58	11.8	295	5	PCT-US95-12675-1 Sequence 1, Appl
45	58	11.8	770	1	US-08-445-135-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-740-223A-11  
Sequence 11, Application US/08740223A  
Patent No. 6265564  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: Expressed Ligand - Vascular  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneon Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,223A  
FILING DATE: 25-OCT-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 60/022/999  
FILING DATE: 02-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Coberl, Robert J  
REGISTRATION NUMBER: 36,108  
REFERENCE/DOCKET NUMBER: REG 333  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: MTL3  
LOCATION: 1...503  
OTHER INFORMATION: mouse TIE ligand-3  
US-08-740-223A-11  
Query Match 13.8%; Score 68; DB 4; Length 503;

Best Local Similarity 28.48; Pred. No. 1;  
Matches 21; Conservative 18; Mismatches 29; Indels 6; Gaps 2.

```
Oy      27 VMETIHYNTAMMRKSLSEASTVREKVALLICBELQSIINGSGGEAMVLVEYWRSLSAEG   86
```

---

```
Dd      159 VLNDGTLMHKQMLENSTL-STNKLEPROMIQSRBLQRLOGRN---RALETRRLDALEAQN    212
```

```

0Y      87  QAHVLAIRDVEEQI 100
      || : :| :| :| :
Db      213 QAQLNSLQEKREQL 226

```

RESULT 2  
US-08-665-926-8

Patent No. 5851797  
GENERAL INFORMATION:  
APPLICANT: Valenzuela et al.  
TITLE OF INVENTION: THE LIGAND-3, METHODS OF MAKING AND USES

TITLE OF INVENTION: THEREOF  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Regeneron Pharmaceuticals, Inc  
 STREET: 777 Old Saw Mill River Road

```

;      CITY:  Parrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-6707
;      COMPUTER READABLE FORM:

```

```

?
? MEDIUM TYPE: Floppy disk
? COMPUTER : IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/065,926  
 FILING DATE: 19-JUN-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Robert J. Cobert

REGISTRATION NUMBER: 36,108  
REFERENCE/DOCKET NUMBER: REG 330-H  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 345-7400  
TELEFAX: (914) 345-2113

```

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 509 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;

```

```

; MOLECULE TYPE: protein
us-08-665-926-8
Query Match 13

```

Query Match	13.8%;	Score 68;	DB 2;	Length 509;
Best Local Similarity	28.4%;	Pred. NO. 1;		
Matches	21;	Conservative	18;	Mismatches 29;
			Indels	6;
			Gaps	2

```

OY 27 VMETIHYNTAMRASLEASIVREBVALIKELQSIQSGGEMAVLYEWRSLAEG 86
    | : : : | : : : | : : : | : : : | : : : |
DB 165 VLNOITLHMKTQMLENSL-STYKKLEKQMLQSLRELQRLQGRN----RALETRRLQALEAQH 218

```

```
QY      87 QAHLALRDVEQI 100
         ||:::|::||:
Db      219 QAQLNSLQEKREQL 232
```

```

RESULT      3
US-08-740-223A-10
; Sequence 10, Application US/08740223A

```

Patent No. 6265564  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: Expressed Ligand - Vascular

```

; TITLE OF INVENTION: Intercellular Signalling Molecules
;
; NUMBER OF SEQUENCES: 28
;

```

ADDRESSEE: Regeneron Pharmaceuticals, Inc  
STREET: 777 Old Saw Mill Road  
CITY: Tarrytown

```

;
;      .COUNTRY:  USA
;      ZIP:      10591
;      COMPUTER READABLE FORM:
;

```

```

;
; OPERATING SYSTEM:  DOS
; SOFTWARE:  FastSEQ Version 2.0
;

```

APPLICATION NUMBER: US/08/740, 223A  
 FILING DATE: 25-OCT-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: USSN 60/022,999  
FILING DATE: 02-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Cobert, Robert J  
REGISTRATION NUMBER: 36,108

REFERENCE/DOCKET NUMBER: REG 33  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ ID NO: 10:

```

: SEQUENCE CHARACTERISTICS:
:     LENGTH: 509 amino acids
:     TYPE: amino acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:

```

```

; MOLECULE TYPE: prot
; FRAGMENT TYPE: inte
; FEATURE:
; NAME/KEY: TIE lig
; LOCATION: 1..509
;

```

OTHER INFORMATION:  
US-08-740-223A-10

[illegible]

Db 165 VLN0TLMMK7OMLENSL--STINKEROML0SREL0L0GRN----RALET0UA0EA0H 218  
Qy 87 QAHVLAIRDVEEQI 100  
|| : : || :  
219 QAOINSG0ERRE0L 232

Db 219 QAQLNSLOEKRRQL 232

US-08-472-028A-4  
; Sequence 4, Application US/08472028A  
; Patent No. 5767373

APPLICANT: Ward,  
APPLICANT: Volrat  
TITLE OF INVENTION:

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive

```

; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,028A  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James, Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1748/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 508 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-028A-4

Query Match 13.7%; Score 67.5; DB 1; Length 508;  
Best Local Similarity 26.7%; Pred. No. 1.2; Indels 11; Gaps 3;  
Matches 20; Conservative 16; Mismatches 28

OY 15 HHFKSKDELIRAVM-----ETTHNTAMMRAS---LEEASTVEREVLALIRCELOSIM 65  
DB 366 HGFKTLGLTFSSMFPDRSPDVHLTYTFIGSRNDELAKAST--DELKQVYTSDLQRL 423  
OY 66 GSGGEMAVLYEWR 80  
DB 424 GVEGEPVSVNHYWR 438

RESULT 5  
US-08-808-931-4  
Sequence 4, Application US/08808931  
GENERAL INFORMATION:  
APPLICANT: Volrath, Sandra  
APPLICANT: Johnson, Marie  
APPLICANT: Potter, Sharon  
APPLICANT: Ward, Eric  
TITLE OF INVENTION: DNA Molecules Encoding Plant  
TITLE OF INVENTION: Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5939602artis Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,931  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612

FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 508 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-931-4

Query Match 13.7%; Score 67.5; DB 2; Length 508;  
Best Local Similarity 26.7%; Pred. No. 1.2; Indels 11; Gaps 3;  
Matches 20; Conservative 16; Mismatches 28

OY 15 HHFKSKDELIRAVM-----ETTHNTAMMRAS---LEEASTVEREVLALIRCELOSIM 65  
DB 366 HGFKTLGLTFSSMFPDRSPDVHLTYTFIGSRNDELAKAST--DELKQVYTSDLQRL 423  
OY 66 GSGGEMAVLYEWR 80  
DB 424 GVEGEPVSVNHYWR 438

RESULT 6  
US-08-808-323-4  
Sequence 4, Application US/08808323  
GENERAL INFORMATION:  
APPLICANT: Johnson, Marie  
APPLICANT: Volrath, Sandra  
APPLICANT: Ward, Eric  
TITLE OF INVENTION: Promoters from Plant  
TITLE OF INVENTION: Protoporphyrinogen Oxidase Genes  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6018105artis Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,323  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1846

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 508 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-808-323-4

Query Match      13.7% Score 67.5; DB 3; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFSKDEILRAVM-----EETIHNTAMRAS-----LEASTYRERYLALIRCELOSIM 65
Db 366 HGFTLGLTFSSMPEPDRSPDVHLTYTFIGSRNOELAKST--DELKQVTSDLQRL 423
QY 66 GSGSEMAVLVEWR 80
Db 424 GVEGEPVSVNHVWR 438

RESULT 7
US-09-050-603A-4
: Sequence 4, Application US/09050603A
: Patent No. 6023012
: GENERAL INFORMATION:
: APPLICANT: Voltrath, Sandra
: APPLICANT: Johnson, Marie
: APPLICANT: Potter, Sharon
: APPLICANT: Ward, Eric
: APPLICANT: Helfetz, Peter
: TITLE OF INVENTION: DNA Molecules Encoding Plant
: TITLE OF INVENTION: Prototophyllinogen Oxidase
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6023012artls Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/050,603A
: FILING DATE: 30-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/808,931
: FILING DATE: 28-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/012,705
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/013,612
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/020,003
: FILING DATE: 21-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: CGC 1847
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
```

```
INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 508 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-050-603A-4

Query Match      13.7% Score 67.5; DB 3; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFSKDEILRAVM-----EETIHNTAMRAS-----LEASTYRERYLALIRCELOSIM 65
Db 366 HGFTLGLTFSSMPEPDRSPDVHLTYTFIGSRNOELAKST--DELKQVTSDLQRL 423
QY 66 GSGSEMAVLVEWR 80
Db 424 GVEGEPVSVNHVWR 438

RESULT 8
US-09-102-420B-4
: Sequence 4, Application US/09102420B
: Patent No. 6084155
: GENERAL INFORMATION:
: APPLICANT: Voltrath, Sandra
: APPLICANT: Johnson, Marie
: APPLICANT: Ward, Eric
: APPLICANT: Helfetz, Peter
: TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN
: TITLE OF INVENTION: OXIDASE ("PROTOX")
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6084155artls Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/102,420B
: FILING DATE: 22-JUN-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/059,164
: FILING DATE: 13-APR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/050,603
: FILING DATE: 30-MAR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/126,430
: FILING DATE: 11-MAR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/808,931
: FILING DATE: 28-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/012,705
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/013,612
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/020,003
: FILING DATE: 21-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/472,028
```

FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 508 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-102-420B-4

Query Match 13.7%; Score 67.5; DB 3; Length 508;  
Best Local Similarity 26.7%; Pred. No. 1.2;  
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFKSKDELRAVM-----ETIHYNTAMRAS---LEASTVRRVLAIRCELOSLIM 65  
| | | : : : | : | : : | : | : : | : | : :  
DB 366 HGFKTLGLTFSSMPPDRSPDVHLTYTFIGSRNOELAKAST--DELKOVYTSLORL 423  
| | | : : : | : | : : | : | : : | : | : :  
QY 66 GSGEAMAVLYEWR 80  
| | | : : : | : | : : | : | : : | : | : :  
DB 424 GVEGPVSVNHYWR 438

## RESULT 9

US-09-071-296-4

; Sequence 4, Application US/09071296  
; Patent No. 6177245

## ; GENERAL INFORMATION:

; APPLICANT: Ward, Eric R  
; APPLICANT: Volrath, Sandra  
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,296  
; FILING DATE: 06-JUN-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/261,198  
; FILING DATE: 16-JUN-94

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 508 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

; MOLECULE TYPE: protein  
; US-09-071-296-4

Query Match 13.7%; Score 67.5; DB 4; Length 508;  
Best Local Similarity 26.7%; Pred. No. 1.2;  
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFKSKDELRAVM-----ETIHYNTAMRAS---LEASTVRRVLAIRCELOSLIM 65  
| | | : : : | : | : : | : | : : | : | : :  
DB 366 HGFKTLGLTFSSMPPDRSPDVHLTYTFIGSRNOELAKAST--DELKOVYTSLORL 423  
| | | : : : | : | : : | : | : : | : | : :  
QY 66 GSGEAMAVLYEWR 80  
| | | : : : | : | : : | : | : : | : | : :  
DB 424 GVEGPVSVNHYWR 438

## RESULT 10

US-09-196-268-4

; Sequence 4, Application US/09196268  
; Patent No. 6282837

## ; GENERAL INFORMATION:

; APPLICANT: Ward, Eric R  
; APPLICANT: Volrath, Sandra  
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/196,268  
; FILING DATE: 06-JUN-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/261,198  
; FILING DATE: 16-JUN-94

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 508 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-196-268-4

Query Match 13.7%; Score 67.5; DB 4; Length 508;  
Best Local Similarity 26.7%; Pred. No. 1.2;  
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFKSKDELRAVM-----ETIHYNTAMRAS---LEASTVRRVLAIRCELOSLIM 65  
| | | : : : | : | : : | : | : : | : | : :  
DB 366 HGFKTLGLTFSSMPPDRSPDVHLTYTFIGSRNOELAKAST--DELKOVYTSLORL 423  
| | | : : : | : | : : | : | : : | : | : :  
QY 66 GSGEAMAVLYEWR 80  
| | | : : : | : | : : | : | : : | : | : :  
DB 424 GVEGPVSVNHYWR 438

```

1 3
2
3  RESULT 11
4  US-09-015-683-4
5  : Sequence 4, Application US/09015683
6  : Patent No. 628306
7  : GENERAL INFORMATION:
8  : APPLICANT: Ward, Eric R
9  : APPLICANT: Voltrath, Sandra
10 : TITLE OF INVENTION: Manipulation of Protoporphyrinogen
11 : TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
12 : NUMBER OF SEQUENCES: 12
13 : CORRESPONDENCE ADDRESS:
14 : ADDRESSEE: Ciba-Geigy Corporation
15 : STREET: 7 Skyline Drive
16 : CITY: Hawthorne
17 : STATE: NY
18 : COUNTRY: USA
19 : ZIP: 10532
20 : COMPUTER READABLE FORM:
21 :
22 :
23 :
24 :
25 :
26 :
27 :
28 :
29 :
30 :
31 :
32 :
33 :
34 :
35 :
36 :
37 :
38 :
39 :
40 :
41 :
42 :
43 :
44 :
45 :
46 :
47 :
48 :
49 :
50 :
51 :
52 :
53 :
54 :
55 :
56 :
57 :
58 :
59 :
60 :
61 :
62 :
63 :
64 :
65 :
66 :
67 :
68 :
69 :
70 :
71 :
72 :
73 :
74 :
75 :
76 :
77 :
78 :
79 :
80 :
81 :
82 :
83 :
84 :
85 :
86 :
87 :
88 :
89 :
90 :
91 :
92 :
93 :
94 :
95 :
96 :
97 :
98 :
99 :
100:

```

```

      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/191,998
        FILING DATE: 06-JUN-1995
        CLASSIFICATION:
          PRIOR APPLICATION DATA:
            APPLICATION NUMBER: US 08/261,198
            FILING DATE: 16-JUN-94
            ATTORNEY/AGENT INFORMATION:
              NAME: Elmer, James Scott
              REGISTRATION NUMBER: 36,129
            TELECOMMUNICATION INFORMATION:
              TELEPHONE: 919-541-8614
              TELEFAX: 919-541-8689
            INFORMATION FOR SEQ ID NO: 4:
              SEQUENCE CHARACTERISTICS:
                LENGTH: 508 amino acids
                TYPE: amino acid
                TOPOLOGY: linear
              MOLECULE TYPE: protein
            US-09-191-998-4

Query Match           13.7%; Score 67.5; DB 4; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3

OY   15 HHFKSKDELRAVM-----ETIHYNTAMNRS----LEEASTVERVALIRCELOSIM 65
      ||| : : : | : | : | : ||| : : : ||| : :
Db   366 HGFKTLGLFFSSMMPDPSPDVHLTYTFIGSGRNOELAKST--DEIKOVVTSLORL 423
      ||||| : : ||| |||

OY   66 GGSGEAMAVLYVEWR 80
      ||| : : ||| |||

Db   424 GVEGEPVSVNHYWR 438

RESULT 13
US-08-449-933-2
Sequence 2, Application US/08449933
Patent No. 5859195
GENERAL INFORMATION:
APPLICANT: Collins, Francis S.
APPLICANT: Wallace, Margaret R.
APPLICANT: Marchuk, Douglas A.
APPLICANT: Anderson, Lone B.
APPLICANT: Guttmann, David H.
TITLE OF INVENTION: Neurofibromatosis Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,933
FILING DATE: 25-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Konski, Antoinette F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-20553.10
TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141 MRSNFOERS SFO  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2818 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENE:  
CHROMOSOME/SEGMENT: 17q11.2  
FEATURE:  
NAME/KEY: Cleavage-site  
LOCATION: group(583..586, 815..818, 2573..2576, 2810..2813)  
OTHER INFORMATION: /note= "Potential CAMP-dependent  
OTHER INFORMATION: protein kinase recognition sites"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2549..2556  
OTHER INFORMATION: /note= "Potential tyrosine  
OTHER INFORMATION: phosphorylation site"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(1264, 1276, 1358, 1377, 1389, 1390, 1391,  
1395, 1396, 1400, 1423, 1426, 1429, 1430)  
OTHER INFORMATION: /note= "Invariant residues within  
OTHER INFORMATION: most statistically significant regions of similarity among th  
OTHER INFORMATION: GAP family of proteins"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(1264..1290, 1345..1407, 1415..1430)  
OTHER INFORMATION: /note= "Most statistically  
OTHER INFORMATION: significant regions of similarity among the GAP family of  
OTHER INFORMATION: proteins"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 496  
OTHER INFORMATION: /note= "At variance with previously  
OTHER INFORMATION: published sequence which shows an ATG methionine codon rather  
OTHER INFORMATION: than an ATA isoleucine codon"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1183  
OTHER INFORMATION: /note= "At variance with previously  
OTHER INFORMATION: published sequence. Shows an CTG leucine codon rather than th  
OTHER INFORMATION: previously published CTC"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1555  
OTHER INFORMATION: /note= "At variance with previously  
OTHER INFORMATION: published sequence. Lacks an extra CAT histidine condon after  
OTHER INFORMATION: this residue"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (2771-2772)  
OTHER INFORMATION: /note= "Position of an 18 amino  
OTHER INFORMATION: acid insertion(SEQ ID NO:10) representing an alternatively  
OTHER INFORMATION: spliced product"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (1370-1371)  
OTHER INFORMATION: /note= "Position of a 21 amino acid  
OTHER INFORMATION: insertion representing an alternatively spliced product"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1125..1537  
OTHER INFORMATION: /note= "NFI catalytic domain"  
FEATURE:

NAME/KEY: Modified-site  
LOCATION: 2746..2818  
OTHER INFORMATION: /note= "Corresponding amino acids  
OTHER INFORMATION: for the PstI-HindIII fragment designated pMAL.B3A"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 65..371  
OTHER INFORMATION: /note= "Corresponding amino acids  
OTHER INFORMATION: for the HpaI-PstI fragment designated pMAL.HF3A.P"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 65..1240  
OTHER INFORMATION: /note= "Corresponding amino acids  
OTHER INFORMATION: for the HpaI-XhoI fragment designated pMAL.HF3A.X"  
PUBLICATION INFORMATION:  
AUTHORS: Wallace, M.R. et al.  
TITLE: Type 1 Neurofibromatosis Gene: Correction  
JOURNAL: Science  
VOLUME: 250  
ISSUE: 12/21/90  
PAGES: 1749-  
DATE: 12/21-1990  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818  
PUBLICATION INFORMATION:  
AUTHORS: Wallace, M.R. et al.  
TITLE: Type 1 Neurofibromatosis Gene: Identification  
TITLE: of a large transcript in three NFI patients  
JOURNAL: Science  
VOLUME: 249  
ISSUE: 07/13/90  
PAGES: 181-186  
DATE: 07/13-1990  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818  
US-08-449-933-2

Query Match 13.6%; Score 67; DB 2; Length 2818;  
Best Local Similarity 26.5%; Pred. No. 17;  
Matches 26; Conservative 18; Mismatches 32; Indels 22; Gaps 4;

QY 12 STFHFKSKDELIRAVMETIH-----YNTAMRASLEE-----ASTVERVLALIRCL 61  
| | | | | : | | | | | : | | | | | : | | | | | :  
DB 708 SCFRHCEADIRCAVEDEVSHLLPNYTFMEFASVSNMSTGRALOKRYMALLR-RI 766  
QY 62 OSIIMGSGEAMAVLYEW-----RSLASGQA 88  
| | | | | : | | | | | : | | | | | : | | | | | :  
DB 767 EHPYAGNTEAMEDTHAKMEQATKLLINYPKAKMEDQA 804

RESULT 14  
US-07-966-049A-2  
Sequence 2, Application US/07966049A  
Patent No. 623861  
GENERAL INFORMATION:  
APPLICANT: Collins, Francis S.  
APPLICANT: Wallace, Margaret R.  
APPLICANT: Marchuk, Douglas A.  
APPLICANT: Anderson, Lone B.  
APPLICANT: Gultman, David H.  
TITLE OF INVENTION: Neurofibromatosis Gene  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:23:32 ; Search time 19 Seconds  
(without alignments)  
400.919 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIQSGSIFHFKSK.....SLSAEGQAHVLAIRDYEQI 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	429	87.0	198	2	C83286
2	110.5	22.4	190	2	S38906
3	110.5	22.4	197	2	F75281
4	94.5	19.2	200	2	C70604
5	92.5	18.8	194	2	F69985
6	92.5	18.8	215	2	F84019
7	89	18.1	192	2	C70487
8	86	17.4	195	2	F84037
9	85	17.2	197	2	G82973
10	84.5	17.1	215	2	T36800
11	84	17.0	206	2	T35415
12	82.5	16.7	223	2	T35424
13	81	16.4	190	2	D70027
14	80	16.2	186	2	B83491
15	80	16.2	212	2	H83274
16	79	16.0	209	2	T36702
17	77.5	15.7	187	2	T37083
18	77.5	15.7	194	2	A83371
19	77	15.6	198	2	F70328
20	77	15.6	216	2	A81156
21	77	15.6	251	2	T36792
22	76.5	15.5	216	2	B81950
23	76.5	15.5	217	2	T35787
24	76	15.4	195	2	T36378
25	76	15.4	213	2	T36803
26	75.5	15.3	195	2	C85524
27	75.5	15.3	195	2	S10899
28	75.5	15.3	209	2	T34970
29	75	15.2	193	2	D83417

30	74	15.0	188	2	B70736	hypothetical prote
31	74	15.0	225	2	C70500	probable transcrip
32	73.5	14.9	172	2	G84416	hypothetical prote
33	73.5	14.9	210	2	E83471	probable transcrip
34	73.5	14.9	210	2	F83393	probable transcrip
35	73	14.8	176	2	C55208	socA3 protein - My
36	72.5	14.7	186	2	C83042	probable transcrip
37	72.5	14.7	213	2	F70945	probable regulato
38	72	14.6	197	2	A83694	hypothetical prote
39	72	14.6	204	2	F69105	conserved hypotet
40	72	14.6	213	2	F82357	transcription regu
41	71.5	14.5	199	2	B85772	hypothetical prote
42	71.5	14.5	201	2	G70849	probable transcrip
43	71.5	14.5	232	2	A57507	virginiae butanol
44	71	14.4	185	2	D83703	transcription regu
45	71	14.4	200	2	E82203	transcription regu

## ALIGNMENTS

RESULT 1  
C83286  
probable transcription regulator PA2885 [Imported] - Pseudomonas aeruginosa (strain P  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83286  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latig, K.; L  
.; Lofy, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: AB2950; MUID:20437337  
A:Accession: C83286  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <STO>  
A:Cross-references: GB:AE004714; GB:AE004091; NID:g9948965; PIDN:AA06273.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2885

Query Match 87.0% Score 429; DB 2; Length 198;  
Best Local Similarity 84.0%; Pred. No. 1,8e-38;  
Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 DLASAVGIQSGSIFHFKSKDEILRAVMEETIHYNTAMKASLEASTYRERYLALRCE 60  
|||||  
Db 40 DLASAVGIQSGSIFHFKSKDEILRAVMEETIHYNTAMKASLEASTYRERYLALRCE 99  
|||||

QY 61 LOSIMGSGEAMAVLYEWRSLSAEGQAHVLAIRDYEQI 100  
|||||  
Db 100 LOSIMGSGEAMAVLYEWRSLSAEGQAHVLAIRDYEQI 139  
|||||

RESULT 2  
S38906  
hypothetical protein 4 - Clostridium pasteurianum  
C:Species: Clostridium pasteurianum  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
R:Meyer, J.  
submitted to the EMBL Data Library, November 1993  
A:Description: Sequence of 6764 bp EcoRI-Sau3A fragment of Clostridium pasteurianum g  
ry M62754).  
A:Reference number: S38906  
A:Accession: S38906  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-190 <MEY>  
A:Cross-references: EMBL:Z28353; NID:g431946; PIDN:CAA8221.1; PID:g431950  
C:Superfamily: Bacillus subtilis probable transcription regulator yrhI



Query Match	17.4%;	Score 86;	DB 2;	Length 195;
Best Local Similarity	23.8%;	Pred. No. 0.067;		

```
Qy 1 DLASAVGIGSCGFHFHFKSDELLRAVMEETIHYNTAMMASLEASTVRER---VLAL 56
    |:| |: :: |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 46 D1GRAGMSPALYIHYKTKPELLRISDIGHTRAVAILPSAOGGSAELRLADAVSSF 105
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:27:17 ; Search time 10.34 Seconds

(without alignments)  
354.592 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIGSGSIFHFHRSK.....SLSAEQAVLALRDVYEQI 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.5	16.5	203	1	BETI_RHIME
2	75.5	15.3	195	1	BETI_ECOLI
3	74	15.0	188	1	YY05_MYCTU
4	72	14.6	2090	1	HFC1_MESAU
5	70.5	14.3	215	1	YIJC_ECOLI
6	70	14.2	187	1	H893_ECOLI
7	70	14.2	676	1	HS7C_TRYB
8	69.5	14.1	452	1	NH34_CAEBL
9	68.5	13.9	267	1	Y617_SYNY3
10	68.5	13.9	289	1	YERO_BACSU
11	68	13.8	391	1	IF43_NICPL
12	68	13.8	4568	1	DYHC_CAEBL
13	67.5	13.7	226	1	TCMR_STRGA
14	67.5	13.7	1814	1	TSC2_MOUSE
15	67	13.6	210	1	MTRR_NEIGO
16	67	13.6	215	1	ACRR_ECOLI
17	66.5	13.5	199	1	YDHM_ECOLI
18	66.5	13.5	917	1	Y6J3_YEAST
19	66	13.4	218	1	TKR_HAEIN
20	66	13.4	223	1	YWCC_BACSU
21	65.5	13.3	196	1	UIDR_ECOLI
22	65.5	13.3	949	1	GLND_RHIME
23	65.5	13.3	1809	1	TSC2_RAT
24	65	13.2	198	1	TKR_ECOLI
25	65	13.2	1529	1	PDRE_YEAST
26	64.5	13.1	471	1	VG20_BPP22
27	64	13.0	275	1	PORI_PEA
28	64	13.0	3746	1	ACVS_PENCH
29	64	13.0	3791	1	ACVT_PENCH
30	63.5	12.9	1807	1	TSC2_HUMAN
31	63	12.8	338	1	LAMP_CHICK
32	63	12.8	885	1	MVP_RAT
33	63	12.8	2820	1	NFI_RAT

# ALIGNMENTS

34	63	12.8	2839	1	NFI_HUMAN	P21359	homo sapien
35	62.5	12.7	455	1	VPN_HAEIN	P71389	haemophilus
36	62	12.6	174	1	NADM_ARCFU	O27968	archaeoglob
37	62	12.6	219	1	TER3_ECOLI	P03093	escherichia
38	62	12.6	358	1	HRCA_CAUCR	P54303	caulobacter
39	62	12.6	428	1	GFAP_BOVIN	Q28115	bos taurus
40	62	12.6	437	1	MURD_HAEIN	P45063	haemophilus
41	62	12.6	461	1	VIME_ONCMY	P48674	oncorhynchus
42	62	12.6	558	1	ALKT_PSEOL	O00593	pseudomonas
43	61.5	12.5	262	1	OCCP_AGRU	P53117	agrobacteri
44	61.5	12.5	3255	1	POLG_LMWE	P89876	1 genome po
45	61	12.4	210	1	YCFQ_ECOLI	P75952	escherichia

RESULT 1  
BETI\_RHIME STANDARD: PRT: 203 AA.  
AC 069786: 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE REGULATORY PROTEIN BETI.  
CN BETI.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=102F34;  
RX MEDLINE=96409668; PubMed=9736747;  
RA Oesteras M., Boncompagni E., Vincent N., Poggi M.-C., Le Rudulier D.;  
RT "Presence of a gene encoding choline sulfatase in Sinorhizobium  
RT meliloti bet operon: choline-O-sulfate is metabolized into glycine  
RT betaine.";  
RT Proc. Natl. Acad. Sci. U.S.A. 95:11394-11399(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RA Gallibert F., Capela D., Hubler-Barloy F., Gattus M., Batut J.,  
RA Boislard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,  
RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetalle D.,  
RA Vandewol M., Puehler A., Becker A., Weidner S.;  
RL Submitted (MAR-2000) to the SWISS-PROT data bank.  
CC -!- FUNCTION: REPRESSOR INVOLVED IN CHOLINE REGULATION OF THE BET  
CC GENES (BY SIMILARITY).  
CC -!- PATHWAY: OSMOREGULATORY CHOLINE-GLYCINE BETAINIC PATHWAY.  
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL: U39940: AAC13370.1; -;  
DR InterPro: IPR001647; HTH\_Tetr.  
DR Pfam: PF00440; tetr. 1.  
DR PROSITE: PS01081; HTH\_TETR\_FAMILY; FALSE\_NEG.  
KW Transcription regulation; DNA-binding; Repressor; Trans-acting factor.  
FT DNA\_BIND 31 50 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 203 AA: 22039 MW: 749F140947339A13 CRC64:

Query Match 16.5%; Score 81.5; DB 1; Length 203;  
Best Local Similarity 26.4%; Pred. No. 0.11;  
Matches 23; Conservative 18; Mismatches 41; Indels 5; Gaps 1;

```

OY      1 DLASVAGSIOSSIFHHFHSKDEILAAWMEETHYNTAMTARSLEASTVREVALNICE 60
        :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      34 EIAIRACGSPALAHNYFFCSKEOLLATITRSLGLRKDVAVAMKAATIPREVXSLIKVS 93

OY      61 LOSINGSGSEAMAAYLVEMRSLSAEGC 87
        :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db      94 FR----ADQAFAPETVAAMLAFYSEAQ 115

RESULT   2
BETI_ECOLI STANDARD; PRT; 195 AA.
ID BETI_ECOLI AC PI7446;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE REGULATORY PROTEIN BETI.
GN BETI OR B0313.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92065800; PubMed=1956285;
RA Lamarck T., Kaasen E., Esnoc M.W., Falkenberg P., McDougall J.,
RA Strom A.R.;
RT "DNA sequence and analysis of the bet genes encoding the
RT osmoregulatory choline-glycine betaine pathway of Escherichia coli.";
RL Mol. Microbiol. 5:1049-1064(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC1655;
RX MEDLINE=97426617; PubMed=9278603;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M.C., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Durcan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federapoli N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Mamath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REPRESSOR INVOLVED IN CHOLINE REGULATION OF THE BET
CC GENES.
CC -1- PATHWAY: OSMOREGULATORY CHOLINE-GLYCINE BETAINE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52905; CAAB37091.1; -.
DR EMBL; AE000138; AAC73416.1; -.
DR EMBL; U73857; ABAB18039.1; ALT_INT.
DR PIR; S10899; S10899.
DR PIR; S15180; S15180.
DR Ecogen; EG1011; BetI.
DR InterPro; IPR001647; HMH_Tetr.
DR Pfam; PF00440; tetr. 1.
DR PROSITE; PS01081; HMH_TETR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Repressor; Trans-acting factor;
KW Complete proteome.
```

```

FT   DNA_BIND      31          50          H-T-H MOTIF (BY SIMILARITY) .
SQ   SEQUENCE     195 AA;  21815 MW;  B52A7V7B3A605E354 CRC64;
Query Match             15.3%; Score 75.5; DB 1; Length 195;
Best Local Similarity   28.6%; Pred. No. 0.43;
Matches    26; Conservative 14; Mismatches 30; Indels 21; Gaps 4;

OY      2 LASAVGIQSGSPHFHKSKDELIRAVMEETIHYNTAMMRASLEASTREHYLAL---I 57
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB      35 IARRAGVSTGIIISHYFRKRNGLILEATMDT-----TSOLR-----DAVLNRLHALPQSA 84
       :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      58 RCELOISMG-----SGEAAVLYEWRS 81
       ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB      85 EORLOAIVGGNFEDTOVSAAAKVALAEWAS 115
       :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT      3
YY05_MYCTU STANDARD; PRT; 188 AA.
AC      Q50720;
DT      01-NOV-1997 (Rel. 35, Created)
DR      01-NOV-1997 (Rel. 35, Last sequence update)
DE      20-AUG-2001 (Rel. 40, Last annotation update)
DN      HYPOTHEICAL 20.6 KDA PROTEIN RV3405C.
DE      RV3405C OR MT3513 OR MTCY78.23.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Acidimicrobiales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsbly T., Jagiels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RL      Nature 393:537-544(1998).
[2]
RN      SEQUENCE FROM N.A.
RP      STRAIN=CDC 1551 / Oshkosh;
RA      Fleischmann R.D., Allard D., Eisen J.A., Carpenter C., White O.,
RA      Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA      Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Knout H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outposton
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; Z77165; CAB01018.1; -.
DR      EMBL; AE007157; AAK47851.1; -.
DR      TIGR; MT3513; -.
DR      Tuberculist; RV3405C; -.
DR      InterPro; IPRO01647; HTH_Tetr.
DR      Pfam; PF00440; tetr_1.
DR      PRINTS; PR00455; HTHHTTR.
KW      Hypothetical protein; Complete proteome.

```



Page 3

```

FT REPEAT      1072    1097       HCF REPEAT 2.
FT REPEAT      1101    1126       HCF REPEAT 3.
FT REPEAT      1157    1182       HCF REPEAT 4 (DEGENERATE).
FT REPEAT      1295    1320       HCF REPEAT 5.
FT REPEAT      1323    1348       HCF REPEAT 6.
FT REPEAT      1358    1383       HCF REPEAT 7 (DEGENERATE).
FT REPEAT      1423    1448       HCF REPEAT 8.
SQ SEQUENCE     2090 AA; 214942 MW; E495EBB1F2385E17 CRC64;

Query Match          14.6%; Score 72; DB 1; Length 2090;
Best Local Similarity 35.5%; Pred. No. 16;
Matches 22; Conservative 10; Mismatches 20; Indels 10; Gaps 2;

OY      36 TAMMARSLEPASTVREVRLLA---IRCLQSLQGSGGEM-----AVLYEMRSLSAE 85
Db      1667 TAAEAFAAAQAAATFEAPQLAIQAVLQAQQAQMVGTEPMPTSEMAAAYTQAEIGHSAE 1726
        |||.:|.:.|||.:|.:.|||.:|.:.|||.:|.:.|||.:|.:.|||.:|.:.|||.:|.:.|||
OY      86 GO 87
Db      1727 GQ 1728

RESULT 5
YJJC_ECOLI STANDARD; PRT; 215 AA.
ID YJJC_ECOLI
AC P27307;
DT 01-AUG-1992 (Rel. 23, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update),
DE HYPOTHEICAL 24.4 KDA PROTEIN IN UDHA-TDMA INTERGENIC REGION (ORF).
DE YJJC OR B3963.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94089392; PubMed=826557;
RA Blotter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
   Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RL region from 89.2 to 92.8 minutes.";
RL Nucleic Acids Res. 21:5408-5417(1993).
RM [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=93077482; PubMed=1447162;
RA Gustafsson C., Warne S.R.;
RT "Physical map of the oxyR-trmA region (minute 89.3) of the
RL Escherichia coli chromosome.";
RL Bacteriol. 174:7878-7879(1992).
CC -I SIMILARITY: STRONG, TO H.INFLUENZAE H10570.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensesh@isb-sib.ch).
CC -----
DR EMBL; U00006; AAC43069.1; ALT_INT.
DR EMBL; AE000470; AAC76945.1; ALT_INIT.
DR EMBL; X66026; CAA46823.1; ALT_INIT.
DR PIR; S21564; S21564.
DR Ecogene; EG11394; yJJC.
DR InterPro; IPR001647; HTH_Tetr.
DR Pfam; PF00440; tetr_2.
DR Hypothetical protein; Complete proteome.
FW CONFLICT 28 S->T (IN REF. 2).
FT CONFLICT 42 G->V (IN REF. 2).
```



OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditoidea;  
OC Rhabditidae: Peloderinae: Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Lloyd C.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ORPHAN NUCLEAR RECEPTOR.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z68217; CA92467.1; -  
DR HSSP: P20393; 1A6Y.  
DR Wormpep: F5866.5; CE03432.  
DR InterPro: IPR000536; Hormone\_rec\_1lg.  
DR InterPro: IPR001628; zf-C4.  
DR Pfam: PF00104; hormone\_rec.1.  
DR Pfam: PF00105; zf-C4; 1.  
DR PRINTS: PR00047; STROIDFINGER.  
DR SMART: SM00430; HOL1; 1.  
DR SMART: SM00399; znf\_C4; 1.  
DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor: Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger.  
FT DNA\_BIND 19 84 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 19 39 C4-TYPE.  
FT ZN\_FING 55 79 C4-TYPE.  
SO SEQUENCE 452 AA; 51885 MW; 1A08677726C391CF CRC64;  
  
Query Match 14.1%; Score 69.5; DB 1; Length 452;  
Best Local Similarity 26.1%; Pred. No. 5;  
Matches 24; Conservative 12; Mismatches 31; Indels 25; Gaps 3;  
  
OY 6 VGIOGSIFFHFKSKDELIRAVMEETIHYNTAMNRASLEASTYREVALIRCLQSIIM 65  
DB 301 VALNGRCNHFRIISEHL--MEDLV--IPREMDMEGEFVLKALILFRAH----- 349  
OY 66 GSGEAMAVLYEWRSLSAEGCAHVLAIRDYV 97  
DB 350 -----RRISEGRRAHKVRDKY 367  
  
RESULT 9  
Y617-SYNY3 STANDARD: PRT: 267 AA.  
AC 055707;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 28.9 KDA PROTEIN SLI0617.  
GN SLI0617.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiura M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome."  
RT DNA Res. 2:153-166(1995).

CC -!- SIMILARITY: BELONGS TO THE PSPA/IM30 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D64002; BA010345.1; -  
DR Hypothetical protein; Coiled coil; Complete proteome.  
FT DOMAIN 26 156 COILED COIL (POTENTIAL).  
SO SEQUENCE 267 AA; 28905 MW; 5DDE309FFB0FF1A6 CRC64;  
  
Query Match 13.9%; Score 68.5; DB 1; Length 267;  
Best Local Similarity 30.4%; Pred. No. 3.5;  
Matches 17; Conservative 9; Mismatches 25; Indels 5; Gaps 1;  
  
OY 28 MEETIHYNTAMNRASLEASTYREVALIR-----CELOSTMGSGEAMAVLYE 78  
DB 114 MSENLRNMLAEAKISEAKTKKMLQARAKAKANAELQDTLGIGTSATSATSE 169  
  
RESULT 10  
YERO\_BACSU STANDARD: PRT: 289 AA.  
ID YERO\_BACSU  
AC 031500;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR YERO.  
GN YERO.  
OS Bacillus subtilis.  
OC Bacteria: Firmicutes: Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z99107; CAB12490.1; -  
DR Subtilist: BG12841; yero.  
DR InterPro: IPR001647; HTH\_Tetr.  
DR Pfam: PF00440; tetr.1.  
DR PRINTS: PR00455; HTHTEPR.  
DR PROSITE: PS01081; HTH\_TETR\_FAMILY; 1.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
KW Complete proteome.  
FT DNA\_BIND 25 44 H-T-H MOTIF (POTENTIAL).  
SO SEQUENCE 289 AA; 32846 MW; 540697223084DCBB CRC64;  
  
Query Match 13.9%; Score 68.5; DB 1; Length 289;  
Best Local Similarity 35.6%; Pred. No. 3.8;  
Matches 16; Conservative 11; Mismatches 17; Indels 1; Gaps 1;  
  
OY 1 DLASAVGIQSGSIFFHFKSKDELIRAVMEETIHYNTAMNRASLEAE 45  
DB 28 ELASEGTSKGAFLYLFHFKSKALLLSACEYIYIGSMNMKK-NIEE 71

```
RESULT 11
ID IF43_NICPL STANDARD; PRT; 391 AA.
AC P41380;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE EUKARYOTIC INITIATION FACTOR 4A-3 (EIF4A-3).
OS Nicotiana glauca (leadwort-leaved tobacco).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
OC Asteridae: euasterids I: Solanales: Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RX MEDLINE=92051287; PubMed=1719476;
RA Owttrim G.W., Hofmann S., Kuhlmeier C.;
RT "Divergent genes for translation initiation factor eIF-4A are
RT coordinately expressed in tobacco.";
RL Nucleic Acids Res. 19:5491-5496(1991).
CC -1- FUNCTION: EIF-4A IS BOTH A SUBUNIT OF A HIGH MOLECULAR WEIGHT
CC PROTEIN COMPLEX INVOLVED IN CAP RECOGNITION AND IS REQUIRED AS A
CC SINGLE POLYPEPTIDE CHAIN FOR MRNA BINDING TO RIBOSOME. IT IS AN
CC ATP-DEPENDENT SINGLE STRANDED DNA-BINDING PROTEIN WITH A SEQUENCE-
CC INDEPENDENT UNWINDING ACTIVITY (HELICASE).
CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X61206; CAA43514.1; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000629; DEAD_ATP_helcse.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD; 1. Helicase_C.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00039; DEAD_ATP_HELICASE; FALSE_NEG.
KM Initiation factor: Protein biosynthesis: ATP-binding; RNA-binding;
KM DNA-binding; Helicase; Multigene family.
FT NP_BIND 62 69 ATP (BY SIMILARITY).
FT SITE 167 170 DEAD BOX.
SQ SEQUENCE 391 AA; 44194 MW; 92CD2721916BCB1E CRC64;

Query Match 13.8%; Score 68; DB 1; Length 391;
Best Local Similarity 20.8%; Pred. No. 6.1;
Matches 21; Conservative 27; Mismatches 39; Indels 14; Gaps 2;

OY 7 GIOSGIFHHFKSKDELRAVMEETIHYNTAMMRASL-----DEASTVRRRL 54
DB 13 GVEPIASFAMGIRKDLLRGVYQGEFKPSAIDQRAVLPIISGRDVAQAQSGTGKTSMI 72
OY 55 ALIRCELOSTMGSGEAMAVLYEWRSLAEGQAHVLAIRD 95
DB 73 ALTVCOI--VDTKSEVOALILSPTRLAQTEKVLAIAD 111

RESULT 12
ID DYHC_CAEEL STANDARD; PRT; 4568 AA.
AC 019020;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

```
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
OS Caenorhabditis elegans.
OC Eukaryota: Metazoa: Nematoea: Chromadorea: Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=96114101; PubMed=8674131;
RA Lye R.J., Wilson R.K., Waterston R.H.;
RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
RT nematode Caenorhabditis elegans.";
RL Cell Motil. Cytoskeleton 32:26-36(1995).
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L33260; AAC37251.1; -.
DR InterPro: IPR003593; AAA.
DR SMART: SM00382; AAA; 1.
KM Motor protein: Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 587 652 COILED COIL (POTENTIAL).
FT DOMAIN 814 844 COILED COIL (POTENTIAL).
FT DOMAIN 1241 1274 COILED COIL (POTENTIAL).
FT DOMAIN 1324 1340 COILED COIL (POTENTIAL).
FT DOMAIN 1559 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1964 1992 COILED COIL (POTENTIAL).
FT DOMAIN 3132 3229 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3339 3432 COILED COIL (POTENTIAL).
FT DOMAIN 3707 3739 COILED COIL (POTENTIAL).
FT DOMAIN 4359 4386 COILED COIL (POTENTIAL).
FT NP_BIND 1865 1872 ATP (POTENTIAL).
FT NP_BIND 2163 2170 ATP (POTENTIAL).
FT NP_BIND 2537 2544 ATP (POTENTIAL).
FT NP_BIND 2880 2887 ATP (POTENTIAL).
SQ SEQUENCE 4568 AA; 521568 MW; 028E52684F381676 CRC64;

Query Match 13.8%; Score 68; DB 1; Length 4568;
Best Local Similarity 27.7%; Pred. No. 1.1e+02;
Matches 26; Conservative 9; Mismatches 19; Indels 40; Gaps 4;

OY 32 IHVNTAMMRASLEASTVRRERVALIRCELOSTMG----- 66
DB 2106 IHVTANOMRELRLOOLSTVCEHL-LIYSDVQEGSGMWLDPKVLQVLTNLMHGLMVG 2164
OY 67 -GSGEAMAVLYEWRSL-----SAEGQAHV 91
DB 2165 SSGSGKTMA-----WKVLLKALERWENEGVAHY 2193

RESULT 13
ID TCMR_STRGA STANDARD; PRT; 226 AA.
AC P39885;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TETRACENOMYCIN C TRANSCRIPTIONAL REPRESSOR.
GN TCMR.
```

OS Streptomyces glaucescens.  
 CC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
 CC Actinomycetales: Streptomyces: Streptomyces: Streptomyces.  
 CC NCBI\_TaxID=1907;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GLA.0;  
 RX MEDLINE=92276347; PubMed=1592819;  
 RA Guilfoile P.G., Hutchinson C.R.;  
 RT "Sequence and transcriptional analysis of the Streptomyces  
 RT glaucescens tcmr tetracycline C resistance and repressor gene  
 RT loc1."  
 RL J. Bacteriol. 174:3651-3658(1992).  
 [2]  
 RP FUNCTION.  
 RX MEDLINE=92276348; PubMed=1592820;  
 RA Guilfoile P.G., Hutchinson C.R.;  
 RT "The Streptomyces glaucescens tcmr protein represses transcription of  
 RT the divergently oriented tcmr and tcmr genes by binding to an  
 RT intergenic operator region."  
 RL J. Bacteriol. 174:3659-3666(1992).  
 CC -1- FUNCTION: REPRESSES TRANSCRIPTION OF THE DIVERGENTLY ORIENTED TCMR  
 CC AND TCMR (TETRACYCLINE C RESISTANCE AND EXPORT) GENES BY BINDING  
 CC TO AN INTERGENIC OPERATOR REGION. THIS BINDING IS INHIBITED BY  
 CC TETRACYCLINE C.  
 CC -1- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC TETRACYCLINE  
 CC C.  
 CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M80674; AAA67508.1; ALT\_SEQ.  
 DR PIR: B41901; B41901.  
 DR PIR: S27686; S27686.  
 DR InterPro: IPR001647; HTH\_Tetr.  
 DR Pfam: PF00440; tetr.1.  
 DR PRINTS: PR00455; HTH\_Tetr.  
 DR PROSITE: PS01081; HTH\_Tetr.FAMILY.1.  
 DR Antibiotic biosynthesis; Transcription regulation; DNA-binding;  
 KW Repressor.  
 FT DNA\_BIND 49 68 H-T-H MOTIF (POTENTIAL).  
 FT SEQUENCE 226 AA; 25376 MW; 389C426A8054060C CRC64;  
 Query Match 13.7%; Score 67.5; DB 1; Length 226;  
 Best Local Similarity 28.0%; Pred. No. 3.6;  
 Matches 33; Conservative 16; Mismatches 38; Indels 31; Gaps 6;  
 OY 2 LASAVGIGSGSFHFKSKDEI-----LRVMEETHYNTAMRA-----SLEEST 48  
 DB 53 IAEAVEVHRTFFRHFASKEEVALPISIDE-----AFLAALEVRAGENPLQAMSG 105  
 OY 49 VREVALAIR-CELOSIMGSGEAMAVLV-----YEMSLAEGO-AHVALLRD 95  
 DB 106 ARAVAVIGRVDELGVGALHMAAMRLVETPGLLAFLYLRSEMEGRILARIARE 163  
 RESULT 14  
 TSC2\_MOUSE  
 ID TSC2\_MOUSE STANDARD; PRT: 1814 AA.  
 AC Q61037; Q61007; Q61008; P97723; P97724; P97725; P97727; Q9WUF6;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TUBERIN (TUBEROUS SCLEROSIS 2 HOMOLOG PROTEIN).  
 GN TSC2.

OS Mus musculus (Mouse).  
 CC Eumariota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 CC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 CC NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORMS A-G).  
 RC Tissue-Heart;  
 RX MEDLINE=96258425; PubMed=8777431;  
 RA Kim K.K., Pajak L., Wang H., Field L.J.;  
 RT "Cloning, developmental expression, and evidence for alternative  
 RT splicing of the murine tuberous sclerosis (TSC2) gene product."  
 RL Cell. Mol. Biol. Res. 41:515-526(1995).  
 [2]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE=96430093; PubMed=8833243;  
 RA Olsson P.G., Schofield J.N., Edwards Y.H., Frischauf A.M.;  
 RT "Expression and differential splicing of the mouse TSC2 homolog."  
 RL Mamm. Genome 7:212-215(1996).  
 [3]  
 RP SEQUENCE OF 1-199 FROM N.A.  
 RC STRAIN-BAB/C; TISSUE=Leukocyte;  
 RX MEDLINE=96417643; PubMed=9743625;  
 RA Sarker A.H., Ikeda S., Nakano H., Terato H., Ide H., Imai K.,  
 RA Akiyama K., Tsutsui K., Bo Z., Kubo K., Yamamoto K., Yasui A.,  
 RA Yoshida M.C., Seki S.;  
 RT "Cloning and characterization of a mouse homologue (mthl1) of  
 RT Escherichia coli endonuclease III."  
 RL J. Mol. Biol. 282:761-774(1998).  
 [4]  
 RP SEQUENCE OF 119-1805 FROM N.A.  
 RX MEDLINE=20051947; PubMed=10584558;  
 RA Kleymenova E.V., Declue J.E., Walker C.L.;  
 RT "Genetic variants of the tuberous sclerosis 2 tumour suppressor gene  
 RT in mouse t haplotypes."  
 RL Genet. Res. 74:139-144(1999).  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR. MAY HAVE A FUNCTION IN  
 CC VESICULAR TRANSPORT, BUT MAY ALSO PLAY A ROLE IN THE REGULATION OF  
 CC CELL GROWTH ARREST AND IN THE REGULATION OF TRANSCRIPTION MEDIATED  
 CC BY STEROID RECEPTORS. INTERACTION BETWEEN HAMARTIN AND TUBERIN MAY  
 CC FACILITATE VESICULAR DOCKING. SPECIFICALLY STIMULATES THE  
 CC INTRINSIC GTPASE ACTIVITY OF THE RAS-RELATED PROTEIN RAB1A AND  
 CC RAB5. SUGGESTING A POSSIBLE MECHANISM FOR ITS ROLE IN REGULATING  
 CC CELLULAR GROWTH (BY SIMILARITY).  
 CC -1- SUBUNIT: INTERACTS WITH HAMARTIN. MAY ALSO INTERACT WITH THE  
 CC ADAPTER MOLECULE RABAPIN 5. THE FINAL COMPLEX CONTAINS TUBERIN  
 CC AND RABAPIN 5 LINKED TO RAB5 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. AT STEADY STATE FOUND IN  
 CC ASSOCIATION WITH MEMBRANES (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 7 ISOFORMS: A, B, C, D, E, F AND G (SHOWN  
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC -1- SIMILARITY: CONTAINS 1 RAB/RAN-GAP DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U37775; AAA6902.1; -;  
 DR EMBL: U37775; AAA6901.1; -;  
 DR EMBL: U39818; AAB18754.1; -;  
 DR EMBL: AB009371; BAA28845.1; -;  
 DR EMBL: AF132986; AAD27867.1; -;  
 DR MGD: MGI:102548; Tsc2.  
 DR InterPro: IPR000331; Rap\_GAP.  
 DR InterPro: IPR003913; Tuberin.  
 DR Pfam: PF02145; Rap\_GAP.1.  
 KW Anti-oncogene; Alternative splicing; GTPase activation.  
 FT VARSPPLIC 79 115 MISSING (IN ISOFORM A).  
 FT VARSPPLIC 534 572 MISSING (IN ISOFORM B).

```

FT  VARSPLIC  947  989  MISSING (IN ISOFORM F).
FT  VARSPLIC  947  990  MISSING (IN ISOFORM C).
FT  VARSPLIC  1245 1258  GHAPVIVATATGC -> RDTALYKSLSPVAG (IN
FT  VARSPLIC  1271 1293  ISOFORM E AND ISOFORM F).
FT  VARSPLIC  1693 1728  MISSING (IN ISOFORM D, ISOFORM E AND
FT  VARSPLIC  1693 1728  ISOFORM F).
FT  VARSPLIC  1693 1728  GPACKCEWMPQGEIVVALPVNMLVTTLCHLQ -> M
FT  VARSPLIC  1693 1728  EGLVDTSVARIQSDRNLSFVAROMALHAN (IN ISOFORM
FT  VARSPLIC  1693 1728  E AND ISOFORM F).
FT  VARSPLIC  1776 1814  MISSING (IN ISOFORM F).
FT  VARSPLIC  161 161  MISSING (IN REF. 2).
FT  CONFLICT  269 269  G -> S (IN REF. 2).
FT  CONFLICT  462 462  A -> R (IN REF. 2).
FT  CONFLICT  476 476  I -> N (IN REF. 2).
FT  CONFLICT  549 549  D -> N (IN REF. 2).
FT  CONFLICT  707 707  K -> N (IN REF. 4).
FT  CONFLICT  861 862  VP -> AA (IN REF. 2).
FT  CONFLICT  1127 1127  R -> P (IN REF. 2).
FT  CONFLICT  1665 1665  MISSING (IN REF. 2).
SQ  SEQUENCE  1814 AA: 202069 MW: 913AB87194ADA5B CRC64:

```

```

Query Match 13.7%: Score 67.5; DB 1: Length 1814;
Best Local Similarity 27.5%: Pred. No. 41;
Matches 25; Conservative 17; Mismatches 36; Indels 13; Gaps 3;

```

```

OY 6 VGIOGSIHFHFKSKDEILRAVMEETI-----HYNTAMRASLEASTVREVRALALIR 58
Db 513 VDLAEGCITHFNSLDIEKVMARSLSPELEERDLAVASASLEEDVKTVAGLIVLQ 572
OY 59 CELOSMGSGEAMAVLYEWMSLSAEGQAH 89
Db 573 TKLYTLPA---SHATRYVE--SLISHIOLH 597

```

```

RESULT 15
MTRR_NEIGO STANDARD: PRT: 210 AA.
AC P39897;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE REGULATOR PROTEIN MTRR.
GN MTRR.
OS Neisseria gonorrhoeae.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by
the mtr system.";
RL MOL. MICROBIOL. 11:769-775(1994).
CC -1- FUNCTION: PUTATIVE REPRESSOR OF MTRC GENE. CONTROLS THE
CC PERMEABILITY OF THE CELL ENVELOPE TO HYDROPHOBIC COMPOUNDS SUCH AS
CC ANTI-BIOTICS AND DETERGENTS.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Z25797: CAA81047.1: -
DR PIR: S40250: S40250.
DR HSSP: P09164: 2RT.
DR InterPro: IPR001647: HTH_Tetr.

```

```

DR Pfam: PF00440: tetr. 1.
DR PROSITE: PS01081: HTH_TETR_FAMILY: 1.
KW Transcription regulation; DNA-binding; Repressor.
FT DNA_BIND 32 51 H-T-H MOTIF (POTENTIAL)
FT VARIANT 105 105 H -> Y (IN PERICILLIN-RESISTANT
FT ISOLATES).
SQ SEQUENCE 210 AA: 24192 MW: 41E2646CBFE57F2 CRC64:

```

```

Query Match 13.6%: Score 67; DB 1: Length 210;
Best Local Similarity 16.6%: Pred. No. 3.8;
Matches 24; Conservative 25; Mismatches 32; Indels 64; Gaps 4;

```

```

OY 1 DLASAVGIOGSIHFHFKSKDEILRAVME-----
Db 35 EIAQAGVTRCALYWHKRNKEDLDALFQRIQDDIENCINADADAGSGVTVFRHTLLH 94
OY 30 -----ETHY-----NTAMRASLEASTVREVRALALIR-RCGLQSIM 65
Db 95 FFERLQSDIHFKFHNILFLKCEHTEONAAVIAIARKHOAIWREKITAVALTEAVENQDLA 154
OY 66 GCGSGEAMAV-----LYEWMS 81
Db 155 DDLDEKETAVALFIKSTLDGLIWRMS 179

```

```

Search completed: March 4, 2002, 20:28:54
Job time: 97 sec

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:25:12 ; Search time 27.32 Seconds

(Without alignments)  
535,404 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIQSGSIFHFHFKS.....SLSAEGQAHVLAIRDVEQL 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_17:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	484	98.2	207	2	Q9RF4
2	429	87.0	198	2	Q9H2W2
3	110.5	22.4	190	2	Q59306
4	110.5	22.4	197	2	Q8R4V9
5	94.5	19.2	200	2	P96839
6	92.5	18.8	194	2	P94548
7	92.5	18.8	215	2	Q9K8P5
8	90	18.3	192	2	Q9A8K8
9	90	18.3	222	2	Q9S314
10	89	18.1	192	2	Q67927
11	86	17.4	192	2	Q9K3Q4
12	86	17.4	195	2	Q9K8A4
13	85	17.2	197	2	Q9HTJ0
14	84.5	17.1	215	2	Q9S253
15	84	17.0	206	2	Q9X7X0
16	83	16.8	214	2	Q9JN89
17	82.5	16.7	223	2	Q9X7X9
18	82	16.6	265	2	Q9AA36
19	81	16.4	190	2	Q32228

20	80	16.2	186	2	Q91497	Q91497 pseudomonas
21	80	16.2	212	2	Q9H2P1	Q9H2P1 pseudomonas
22	79	16.0	209	2	Q9XA31	Q9XA31 streptomyc
23	78	15.8	196	2	Q9RM14	Q9RM14 bacillus an
24	78	15.8	211	2	Q9A905	Q9A905 caulobacter
25	77.5	15.8	217	2	Q9K3M6	Q9K3M6 streptomyc
26	77.5	15.7	187	2	Q53165	Q53165 mycobacteri
27	77.5	15.7	194	2	Q911S1	Q911S1 pseudomonas
28	77	15.6	198	2	Q66558	Q66558 aquifex aeo
29	77	15.6	208	2	Q9A905	Q9A905 caulobacter
30	77	15.6	215	2	Q9A690	Q9A690 caulobacter
31	77	15.6	216	2	Q9K017	Q9K017 neisseria m
32	77	15.6	251	2	Q9S261	Q9S261 streptomyc
33	76.5	15.5	202	2	Q9A250	Q9A250 caulobacter
34	76.5	15.5	216	2	Q9J29	Q9J29 neisseria m
35	76.5	15.5	217	2	Q87854	Q87854 streptomyc
36	76	15.4	192	2	Q9Z6B7	Q9Z6B7 streptomyc
37	76	15.4	195	2	Q9X8M4	Q9X8M4 streptomyc
38	76	15.4	202	2	Q9KX78	Q9KX78 streptomyc
39	76	15.4	213	2	Q9S250	Q9S250 streptomyc
40	76	15.4	215	2	Q9AA74	Q9AA74 caulobacter
41	75.5	15.3	209	2	Q9S2V6	Q9S2V6 streptomyc
42	75	15.2	82	2	Q07388	Q07388 mycobacteri
43	75	15.2	190	2	Q9A109	Q9A109 staphylococ
44	75	15.2	193	2	Q91209	Q91209 pseudomonas
45	75	15.2	1489	5	Q9M1R3	Q9M1R3 drosophila

# ALIGNMENTS

## RESULT 1

Q9RF4 ID Q9RF4 PRELIMINARY: PRT: 207 AA.  
AC Q9RF4;  
DT 01-MAY-2000 (TREMUREL\_13, Created)  
DT 01-JUN-2001 (TREMUREL\_17, Last sequence update)  
DT 01-JUN-2001 (TREMUREL\_17, Last annotation update)  
DE PUTATIVE TRANSCRIPTIONAL REGULATOR PF12.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20349410; PubMed=1089151;  
RA Sutton C.L., Kim J., Yamane A., Dalwadi H., Wei B., Landers C.,  
RT Targan S.R., Braun J.;  
RT "Identification of a novel bacterial sequence associated with Crohn's  
disease";  
RL Gastroenterology 119:23-31(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Wei B., Huang T., Dalwadi H., Sutton C.L., Braun J.;  
RT "12, a Crohn's disease-associated microbial gene and T-cell  
superantigen, is a species-specific product of Pseudomonas  
fluorescens";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
DR EMBL: AF173683; AAK16600.1;  
DR InterPro: IPR001647; HTH\_Tetr.  
DR Pfam: PF00440; tetr.1.  
DR PRINTS: PR00455; HTH\_Tetr.  
DR PROSITE: PS01081; HTH\_TETR\_FAMILY.1.  
KW DNA-binding; Transcription regulation.  
SQ SEQUENCE 207 AA; 23107 MW; DD3AFB1FE10A736 CRC64;

Query Match 98.2%; Score 484; DB 2; Length 207;  
Best Local Similarity 98.0%; Pred. No. 2.7e-42;  
Matches 98; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 DLASAVGIQSGSIFHFHFKSKDELLRAVMEETIHYNTAMRASLEFASVREVALIACE 60

```
Db 49 DLSAVGIQSGSIFHHRKSKDELIRAVMEETHTNTAMMRASLEASTVRRVLAIRCK 108
|||
Qy 61 LOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDVEQI 100
|||
Db 109 LOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDVEQI 148

RESULT 2
O9H2M2
ID O9H2M2 PRELIMINARY; PRT; 198 AA.
AC O9H2M2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR.
GN PA2885.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
EMBL: AE004714; AAC06273.1; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; Tetr.
DR PRINTS: PR00455; HTHTEPR.
DR PROSITE: PS01081; HTH_TETR_FAMILY; 1.
KM Complete proteome: DNA-binding; Transcription regulation.
SQ SEQUENCE 198 AA; 22057 MW; 079217CCCTB8FF85 CRC64;

Query Match 87.0%; Score 429; DB 2; Length 198;
Best Local Similarity 84.0%; Pred. No. 1.2e-36;
Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DLSAVGIQSGSIFHHRKSKDELIRAVMEETHTNTAMMRASLEASTVRRVLAIRCE 60
|||
Db 40 DLSAVGIQSGSIFHHRKSKDELIRSVMEETHTNTALMRALADADRLRRVGLIRCE 99

Qy 61 LOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDVEQI 100
|||
Db 100 LOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDVEQI 139

RESULT 3
O59306
ID O59306 PRELIMINARY; PRT; 190 AA.
AC O59306;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S21.
OS Clostridium pasteurianum
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1501;
RN [1]
RP SEQUENCE FROM N.A.
RA Meyer J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
```

```
CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: Z28353; CAA82211.1; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; Tetr. 1.
DR PROSITE: PS01081; HTH_TETR_FAMILY; 1.
KM DNA-binding; Ribosomal protein; Transcription regulation.
SQ SEQUENCE 190 AA; 21692 MW; 0C6ED92A48AA69C0 CRC64;

Query Match 22.4%; Score 110.5; DB 2; Length 190;
Best Local Similarity 27.0%; Pred. No. 0.00085;
Matches 27; Conservative 25; Mismatches 35; Indels 13; Gaps 3;

Qy 1 DLSAVGIQSGSIFHHRKSKDELIRAVMEETHTNTAMMRASLEAS---TVRRVLA 56
|||
Db 28 ELASMGAVAGTLYHHRKSEIFKYTEIGVN---LKMNEIDATDKETLAKAV 83

Qy 57 IRCLOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDV 96
|||
Db 84 CRVQLNLVKNRDFEKVIAQLMGR---ELRLELRDI 118

RESULT 4
O9RRV9
ID O9RRV9 PRELIMINARY; PRT; 197 AA.
AC O9RRV9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR, TETR FAMILY.
GN DR2376.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
EMBL: AE002068; AAF1921.1; -
DR TIGR: DR2376; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; Tetr. 1.
DR PRINTS: PR00455; HTHTEPR.
KM Complete proteome: DNA-binding; Transcription regulation.
SQ SEQUENCE 197 AA; 22307 MW; 769F2960C3F848D CRC64;

Query Match 22.4%; Score 110.5; DB 2; Length 197;
Best Local Similarity 27.5%; Pred. No. 0.00088;
Matches 28; Conservative 23; Mismatches 42; Indels 9; Gaps 2;

Qy 1 DLSAVGIQSGSIFHHRKSKDELIRAVMEETHTNTAMMRASLEASTVRRVLA 56
|||
Db 32 DLAGQLMGGSGLYAHHSKELLVEIRGASQGFDEALSLDYNVNPADKKLRAMFRH 91

Qy 57 IRCLOSIMGSGEAMAVLYEWMRSLSAEGQAHVLAIRDV 98
|||
Db 92 IQ-----VVADNMDSATVFPEHMKHLASBPYAOVVAMRDITD 128

RESULT 5
P96839
```

ID P96839 PRELIMINARY: PRT: 200 AA.  
 AC P96839:  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 22.9 KDA PROTEIN.  
 GN RV357C OR MTCY06G11.04C.  
 OS Mycobacterium tuberculosis.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RC MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 DR EMBL: Z92774; CAB07159.1; -;  
 DR InterPro: IPR001647; HTH\_Tetr.  
 DR Pfam: PF00440; tetr. 1.  
 KW Complete proteome: DNA-binding; Hypothetical protein;  
 KW Transcription regulation.  
 SO SEQUENCE 200 AA; 22906 MW; E1B29D39CF4D955F CRC64;

Query Match 19.2%; Score 94.5; DB 2; Length 200;  
 Best Local Similarity 26.9%; Pred. No. 0.04;  
 Matches 21; Conservative 20; Mismatches 34; Indels 3; Gaps 1;

OY 1 DLASVGIOSGSIFFHFKSKDELRAVMEETIHYNTAMRASLEASTYREVALIRCE 60  
 DB 35 DIADAGILSGSLYHNFASKEMVDELLRGFDLWLFARYRDIVDSTANPLERLQGLFMA 94  
 OY 61 LOSIMGSGEAMAVLYVE 78  
 DB 95 FEALIEHHAQ--VVIYQ 109

RESULT 6  
 P94548 PRELIMINARY: PRT: 194 AA.  
 AC P94548:  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 22.0 KDA PROTEIN.  
 GN YSA.  
 OS Bacillus subtilis.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97217425; PubMed=9063446;  
 RA Goethel S.F., Schmid R., Wipat A., Carter N.M., Emerson P.T.,  
 RA Harwood C.R., Marahiel M.A.;  
 RA "An internal FK506-binding domain is the catalytic core of the prolyl  
 RT isomerase activity associated with the Bacillus subtilis trigger  
 factor."  
 RL Eur J Biochem 244:59-65(1997).  
 SO SEQUENCE FROM N.A.

RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boursier R., Boursier L., Brans A., Braun M., Briggell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Laber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeduchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,  
 RA Tostato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wandut R., Wedler E., Wedler H., Wetzinger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RA "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis."  
 RL Nature 390:249-256(1997).  
 SO SEQUENCE FROM N.A.  
 RP [13]  
 RP STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RA Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 DR EMBL: Z75208; CA99572.1; -;  
 DR EMBL: Z91118; CAB14815.1; -;  
 DR InterPro: IPR001647; HTH\_Tetr.  
 DR Pfam: PF00440; tetr. 1.  
 DR PRINTS: PR00455; HTHTEPR.  
 KW Complete proteome: DNA-binding; Hypothetical protein;  
 KW Transcription regulation.  
 SO SEQUENCE 194 AA; 21978 MW; ABC01C9C3B47CFE8 CRC64;

Query Match 18.8%; Score 92.5; DB 2; Length 194;  
 Best Local Similarity 27.4%; Pred. No. 0.062;  
 Matches 23; Conservative 22; Mismatches 38; Indels 1; Gaps 1;

OY 2 LASAVGIOSGSIFFHFKSKDELRAVMEETIHYNTAMRASLEASTYREVALIRCEL 61  
 DB 32 IAKQGVADGTYLVFKNNEDILISLFKMKGPIERMEDIKKATYKAK-LALVISHK 90  
 OY 62 OSIMGSGEAMAVLYEMRSIAE 85  
 DB 91 FSLAGDHNLAIIVTDELRSNILE 114  
 RESULT 7  
 O9K8P5 PRELIMINARY: PRT: 215 AA.  
 AC O9K8P5:  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE TRANSCRIPTIONAL REGULATOR (TETR/ACRR FAMILY).  
 GN BH2958.  
 OS Bacillus halodurans.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;

```

Query Match 18.3% Score 90: DB 2: Length 192:
Best Local Similarity 25.9% Pred. No. 0.11:
Matches 28: Conservative 22: Mismatches 34: Indels 24: Gaps 4

OY 1 DLASAVGIQSGSIFFHNFSGKDEILAAVMBETHHTYNTAMMRASLEASTVREKVALICE 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32 DVAKAGVAVHNAVHINHTFSGIGVETALDERMKR-----QADRVAVATIOKE 77

```

RESULT	10
067927	
ID	067927
AC	067927;
DT	01-AUG-1998 (TEMBLrel, 07, Created)
DT	01-AUG-1998 (TEMBLrel, 07, last sequence update)
DT	01-JUN-2001 (TEMBLrel, 17, last annotation update)
DE	TRANSCRIPTIONAL REGULATOR (tetr/ACRR FAMILY).
GN	ACRR2 OR AQ_2179.
OS	Aquifex aeolicus.
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.
OX	NCBI_TaxID=63363;
RN	11
RP	SEQUENCE FROM N.A.
RC	STRAIN=VF5;
RX	MEDLINE=98196666; PubMed=9537320;
RA	Decker G., Warren P. V., Gaasterland T., Young M.G., Lenox A.L.,
RA	Graham D.E., Overbeek R., Shead M.A., Keller M., Anjey M., Huber R.,
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT	"The complete genome of the hyperthermophilic bacterium Aquifex
RT	aeolicus.";
RL	Nature 392:353-358(1998).
CC	-1- SIMILARITY: TO THE tetr/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS
DR	EMBL, AE000776; AAC07881.1; -
DR	InterPro: IPR002106; AA_CRM_11gase-II.

DT 01-OCT-2000 (Tremblay, 15, Created)

DR PRINTS: PR00455; HTHETR.  
DR PROSITE: PS01081; HTH\_TETR\_FAMILY; UNKNOWN\_1.

KW Complete proteome; DNA-binding; Transcription regulation.



Tue Mar 5 07:57:19 2002

us-09-966-608-2.rspt

---

